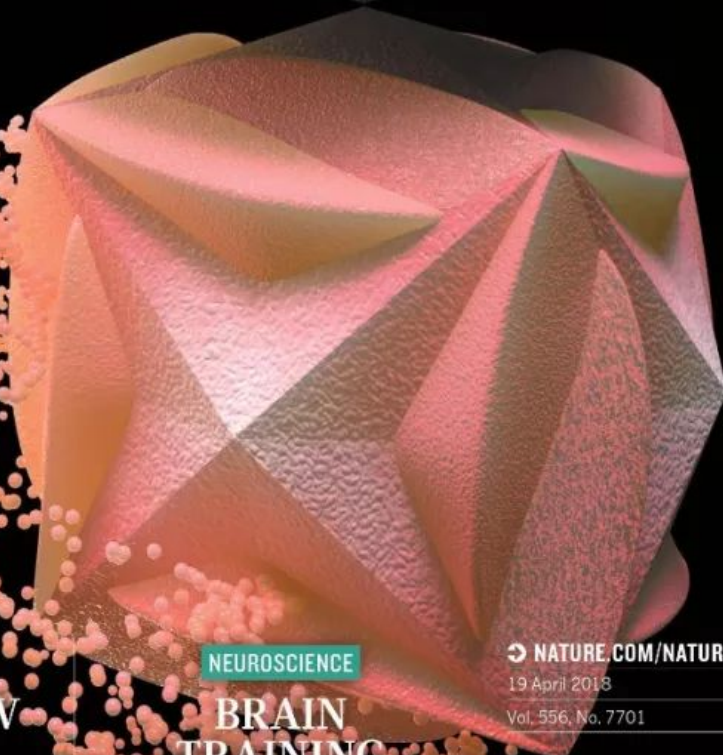
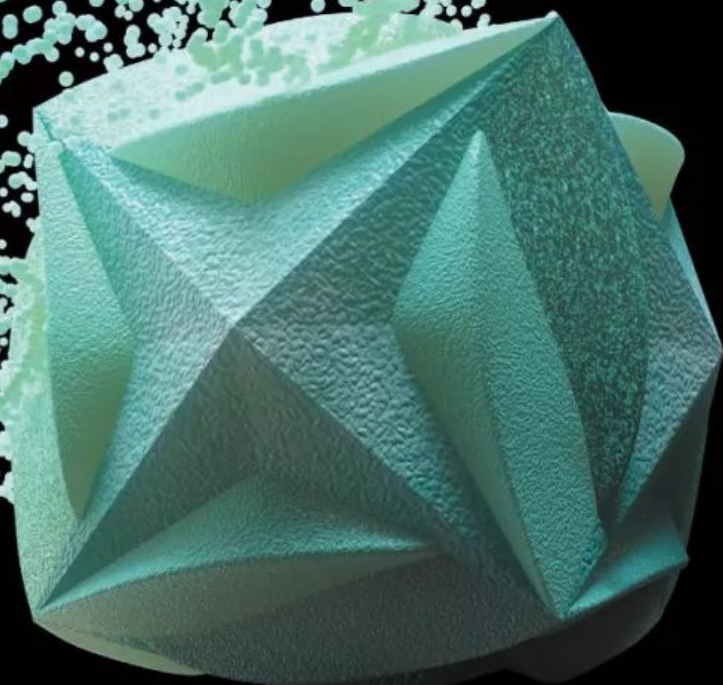


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

PARTICLE REFLECTION

Gold nanostructures gain
chiral properties through
directed growth with
peptides **PAGES 313 & 360**



POLICY

CHEMICAL WEAPONS

Ways to cut the production
of lethal compounds

PAGE 293

SPRING READING

BOOK REVIEW SPECIAL

From the value of value to the
anatomy of conflict

PAGE 300

NEUROSCIENCE

BRAIN TRAINING

Immune cell memories help
slow neurological disorders

PAGES 312 & 332

NATURE.COM/NATURE

19 April 2013

Vol. 556, No. 7701

THIS WEEK

EDITORIALS

AFRICA New guidelines for genomics and other projects **p.274**

WORLD VIEW Improve access for researchers with disabilities **p.275**



BEHAVIOUR Hormone makes birds more likely to share food **p.277**

Military work threatens science

In an uncertain world, more governments are asking universities to help develop weapons. That's a threat to the culture and conscience of researchers.

South Korea is understandably nervous. To the north, a bellicose, belligerent and unpredictable leader has nuclear weapons, increasingly powerful missiles and many troops. South Korea is trying to counter that with technological superiority offered by its robust scientific infrastructure. But the nation's efforts to enhance the technological superiority by using academics to pursue military goals have raised a furor. And South Korea is not the only country to court such controversy.

In February, South Korea opened a centre at its premier research facility, the Korea Advanced Institute of Science and Technology (KAIST) in Daejeon, in collaboration with the country's leading arms manufacturer, Hanwha Systems. Media reports said that the centre, known as the Research Centre for the Convergence of National Defence and Artificial Intelligence, would develop technologies that could be useful for more-advanced weapons, such as missiles that use artificial intelligence (AI) to control their speed and altitude and detect enemy radar in real time.

There was an immediate backlash. Almost 60 AI and robotics researchers from around the world signed an open letter opposing KAIST's participation in an autonomous-weapons race. They threatened to cut all ties with KAIST. But this episode had a happy ending: KAIST's president vowed that the centre wouldn't develop lethal weapons. The boycott was abandoned. This week, the letter's author accepted an invitation to visit KAIST.

But similar fault lines have been exposed elsewhere. Australian scientists continue to debate the government's 2014 defence–science partnerships programme, which has so far enrolled researchers from 32 universities. And a 2016 decision by the European Commission to start funding defence research prompted 400 researchers to sign a petition attacking the move.

In Japan, universities are split over whether they should take funds from the defence ministry's Acquisition, Technology and Logistics Agency. Last year, the advisory board to the nation's cabinet — the Science Council of Japan — called for researchers to boycott the work, and for institutions to set up special committees to evaluate the ethics and propriety of military-related research projects. According to survey results released by the council earlier this month, 46 of the 135 universities polled have such a system in place. But 30 institutions have already allowed researchers to apply, and 41 have no intention of creating such a system. And the nation's astronomical society has voiced support for the fund. It says that its young researchers believe that such work is acceptable if it falls within Japan's policy of maintaining self-defence strategies.

In the United States, university-based military research has long been a fixture, but the push in less-militarized countries points to rising geopolitical uncertainty and instability around the world. Trying to improve defence capabilities in such circumstances is understandable — the issue is where and how it should be done.

More fundamentally, such research threatens core principles that are the bedrock of universities everywhere. A greater reliance on funding for

militarized projects threatens the remit of independent and curiosity-driven research. It breaks down the bonds of trust that connect scientists around the world and undermines the spirit of academic research. The sharing of data and techniques through publications and collaborations has been the basis of peaceful collaborations even between researchers from countries that are at war with each other. If researchers need to question whether their contributions are going to feed development of a weapon, they might — understandably — keep their ideas to themselves.

“The work should align with a fundamental commitment to humane and life-saving applications.”

Government initiatives around the world seem to show that military funds will continue to permeate universities. So be it. But the researchers involved carry a heavy responsibility. The work should align with a fundamental commitment to humane and life-saving applications — drones that can deliver medical supplies to war-torn areas, or robots that can clear minefields, for example. The line is likely to be fuzzy. An AI navigation system seems relatively innocuous for an autonomous surveillance submarine, but in a nuclear submarine, it becomes the kind of application that the global research community protested against in South Korea. Still, as the South Korea example demonstrates, scientists have a crucial role in alerting the world to the potential dangers of emerging technologies, and redirecting the trajectory of the research. Those researchers and institutions that pursue the technologies despite the risks need to remain transparent, so that their peers can not only judge the rigour of their science, but also ensure they steer clear of inhumane applications. ■

Checklist checked

Nature authors say a checklist has improved reproducibility, but more needs to be done.

Five years ago, after extended discussions with the scientific community, *Nature* announced that authors submitting manuscripts to *Nature* journals would need to complete a checklist addressing key factors underlying irreproducibility for reviewers and editors to assess during peer review. The original checklist focused on the life sciences. More recently we have included criteria relevant to other disciplines.

To learn authors' thoughts about reproducibility and the role of checklists, *Nature* sent surveys to 5,375 researchers who had published in a *Nature* journal between July 2016 and March 2017 (see Supplementary information at go.nature.com/2vm2fxw and <https://doi.org/10.6084/m9.figshare.6139937> for the raw data).

Of the 480 who responded, 49% thought that the checklist had

improved the quality of research published in *Nature* (15% disagreed); 37% thought the checklist had improved quality in their field overall (20% disagreed).

Respondents overwhelmingly thought that poor reproducibility is a problem: 86% acknowledged it as a crisis in their field, a rate similar to that found in an earlier survey (*Nature* 533, 452–454; 2016). Two-thirds of respondents cited selective reporting of results as a contributing factor.

Nature's checklist was designed, in part, to make selective reporting more transparent. Authors are asked to state whether experimental findings have been replicated in the laboratory, whether and how they calculated appropriate sample size, when animals or samples were excluded from studies and whether these were randomized into experimental groups and assessed by ‘blinded’ researchers (that is, researchers who did not know which experimental group they were assessing). Of those survey respondents who thought the checklist had improved the quality of research at *Nature* journals, 83% put this down to better reporting of statistics as a result of the checklist.

Is the checklist addressing the core problems that can lead to poor reproducibility? Only partly. Taken as a whole, the responses indicate that we need more nuanced discussions, and more attention on the interconnected issues that result in irreproducibility: training, transparency, publishing pressures and what the report *Fostering Integrity in Research* by the US National Academies of Sciences, Engineering, and Medicine deems “detrimental research practices”.

Journals cannot solve this alone. Indeed, 58% of survey respondents felt that researchers have the greatest capacity to improve the reproducibility of published work, followed by laboratory heads (24%), funders (9%) and publishers (7%).

What role, then, should publishers take? Reproducibility cannot be assessed without transparency, and this is what journals must demand. Readers and reviewers must know how experiments were designed and how measurements were taken and deemed acceptable for analysis; they need to be told about all of the statistical tests and replications.

As such, the checklist (or ‘reporting summary’) provides a convenient tool for revealing the key variables that underlie irreproducibility in an accessible manner for authors, reviewers, editors and readers.

Two studies have compared the quality of reporting in *Nature* journals before and after the checklist was implemented, and with journals that had not implemented checklists. Authors of papers in *Nature* journals are now several times more likely to state explicitly whether they have carried out blinding, randomization and sample-size calculations (S. Han *et al.* *PLoS ONE* 12, e0183591; 2017 and M. R. Macleod *et al.* Preprint at *BioRxiv* <https://doi.org/10.1101/187245>; 2017). Journals without checklists showed no or minimal improvement over the same time period. Even after implementation of the checklist, however, only 16% of papers reported the status of all of the crucial ‘Landis 4’ criteria (blinding, randomization, sample-size calculation and exclusion) for *in vivo* studies — although reporting on individual criteria was significantly higher. Preliminary data suggest that publishing the reporting summaries, as we have done since last year, has resulted in further improvements.

Fortunately, the trend indicated by the survey is positive. Most respondents had submitted more than one paper using the checklist. Nearly half of respondents said they had not considered the checklist until after they had written their first submission; that fell to 31% for subsequent papers, with authors more likely to consider the checklist while planning or performing experiments. Encouragingly, 78% said that they had continued to implement the checklist to some extent, irrespective of their plans to submit to a *Nature* journal in the future.

Progress is slow, but a commitment to enforcement is crucial. That is why we make the checklist and the reporting of specific items mandatory, and monitor compliance. The road to full reproducibility is long and will require perseverance, but we hope that the checklist approach will gain wider uptake in the community. ■

“Respondents overwhelmingly thought that poor reproducibility is a problem.”

Aid from Africa

Africa’s genomics research will benefit from a new set of ethics principles.

Helicopter science. Sample safaris. Parachute research. These are all pejorative terms used to describe the practice of collecting biological samples, artefacts or data from developing countries and analysing them elsewhere, with little input from — or credit given to — local scientists. Such practices are almost universally denounced by research funders and institutions in the global north. Yet the language still crops up, especially in disciplines such as genomics, for which the technology required to decode DNA at high volumes remains concentrated in wealthy countries.

In human genomics, there has been a push to ensure that research on samples collected in developing countries — particularly in Africa — is anchored in local science and community engagement. One example of this is the Human Heredity and Health in Africa (H3Africa) initiative, which is funded by the US National Institutes of Health and the London-based Wellcome Trust. Since 2012, it has funded genomics projects whose principal investigators are African, with several of the projects being managed locally from Kenya’s capital, Nairobi.

As we report this week, the H3Africa group has now published a guide for the ethical handling of genomic research and biobanking in Africa (see <https://doi.org/10.1038/d41586-018-04685-1>). It sets out to empower African researchers and communities, and to educate them on their rights in asking for greater control over how samples are collected, stored and used. It also contains rules of engagement for non-African institutions that are partnering with, or funding research

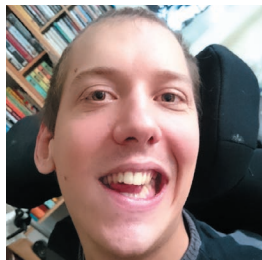
in, Africa. It’s a useful guide, and draws on existing ethics policy documents. Many of its recommendations — such as avoiding tokenistic participation by African researchers, and ensuring that research results are fed back to the communities that donated the samples — have been regarded as good practice in the field for some time. But, in reality, such practices are all too often still lacking.

The fact that the document is derived from in-depth conversations with African researchers and ethics review boards gives it added legitimacy. Perceptions can vary about whether partnerships are equitable or not, and it is not uncommon for northern partners to hold up projects as exemplary in terms of their equity, with African participants in the same projects complaining of limited input. This framework should help, by allowing negotiating partners to sing from the same hymn sheet.

Because it is voluntary, the framework’s impact will depend on its use by its target audiences. African research-ethics committees that preside over applications to carry out genetic research can use it to ensure that their decisions have the interests of Africans at heart. African researchers can draw on it to negotiate more-advantageous terms in partnerships. Research funders can encourage applicants to consider the framework when submitting proposals. African governments can use it to inform their rules guiding genomics research. And, perhaps most importantly, African communities can look to the framework for information about what to expect, or even demand, from their participation in research.

Ultimately, the foremost priority of researchers, funders, regulators and ethicists should be to respect the rights and interests of the populations studied. In the scramble for African genomes, such rights can easily be overlooked — especially in countries with weak governance, where research-ethics rules are outdated or where patient-rights groups are lacking. There is therefore a need for greater involvement by African governments and civil society, to ensure that genomic research is in the public’s interest, not just in the interests of the participating scientists — regardless of where they come from. ■

AARON SCHAAL



Science must rise up to support people like me

Institutions could do more to support researchers who have disabilities, says Aaron Schaal.

The topic I study happens to be similar to one that occupied physicist Stephen Hawking for much of his life. And like he did, I need technology to communicate, as well as round-the-clock personal assistance to live independently; I can control only the movements of my eyes reliably. My family, supervisors and the German health-care system provide essential support. Hawking's death last month renewed focus on accessibility in science. In my experience, there are too many obstacles in academia for people who have physical or mental-health conditions — and who have much to offer to science.

As a 27-year-old doctoral student in mathematical physics, I study why time moves in only one direction. I rely on a very simplistic mathematical model consisting of classical particles that interact through gravity. This should yield ideas for explaining the 'arrow' of time.

Since I was two years old, I have had generalized dystonia — abnormal muscle tone caused by a rare metabolic disease called glutaric aciduria type 1. To communicate and write, I use an eyetracker on my computer or a colour-coded, plexiglass board that I designed myself, which has Latin and Greek letters, as well as numbers and mathematical symbols. Many of my colleagues have learnt to 'listen' to me by tracking which symbols on the board I look at. At first, it can take nearly half an hour to understand a sentence. With practice, the process becomes relatively fast, especially if my conversation partner knows the context. With my PhD adviser, for example, this form of communication takes about three times as long as regular speech.

When interacting with people who can't use the board, I communicate using my eyetracking device, which can produce synthesized voices in different languages, or with the help of personal assistants. The eyetracker allows me to type about two characters per second, assuming that I know exactly what I want to write, that my eyetracker is optimally positioned and calibrated and that the word-prediction software is working well. To give a talk, I formulate the whole thing in full sentences first, so my assistant or laptop can then read it aloud.

During my undergraduate studies, I had to request accommodations, such as extra time for exams. Professors rarely knew what to do: most had had no contact with anyone with a disability. Some couldn't imagine how I would be able to write exams at all.

Things improved as more professors came to know me. The supervisors of both my bachelor's-degree thesis and my master's-degree and PhD work have strongly supported me, which has included dealing with bureaucracy. In my work as a teaching assistant, I have been treated extremely well. I create exercise worksheets, organize tutorials and maintain the website for the course lecture series. These are all things I can do from home, especially if I am unable to go outside. All this shows that individual solutions do exist — if a university is willing to find them.

In 2015, I co-founded Chronically Academic, a global network that connects academics who have disabilities or chronic conditions. Our

website, which I set up and maintain, hosts resources for individuals and institutions (see <https://chronicallyacademic.org>). Last year, we published a series of articles on chronic illness in academia in *The Sociological Review*. Last month, some of us co-organized a conference at University College London called Ableism in Academia. Today, we have some 150 active members offering peer support and raising awareness.

Our experiences show that a research career is possible if you really want it and have support from your family, supervisors and colleagues. But it could be a bit easier. Academics are often expected to move across countries or continents, which can be difficult for some people with certain disabilities, given large variations in health-care systems. Even those who remain in one place must attend conferences. Flying is impossible for me; no commercial aeroplane will transport me in my own

wheelchair. To travel, I need at least two personal assistants and a host of technical and medical equipment. A week-long trip to Tübingen (a 3- to 4-hour car ride from my home in Munich) costs US\$3,000–5,000, including food and lodging for my assistants. I can apply to several German health-care authorities or my university to cover the costs, but that is time-consuming and far from straightforward, and success is not guaranteed.

Once there, events are not always accessible. I have slept on the floor when only bunk beds were available. My colleagues have similar stories. At a conference on inclusion, restrooms were accessible only by stairs, and event planners put planks over the entrance steps for wheelchair access only after an invited scientist refused to be carried in.

One person changed universities after continuous teasing from supervisors and colleagues about involuntary facial movements. Another left when departments refused to even discuss making accommodations.

Needs vary. For me, it's mainly wheelchair access; others need sign-language interpreters. Event planning should include inviting attendees to state any special needs, and working to accommodate them. Institutions should offer and promote training on how to support students and staff. More-flexible contracts — working reduced hours or from a home office — would be a huge improvement. Scholarships and administrative help to cover extra costs of travel and assistance would expand opportunities. And events should be made truly accessible to all participants.

Without these moves, disabilities such as mine will remain out of sight and out of mind in science, making it more homogeneous and less compassionate. That is a severe loss. There are many times when I cannot do anything but think, which lends itself to coming up with new scientific ideas. As Stephen Hawking showed so dramatically, research benefits from a more diverse workforce. ■

Aaron Schaal is a PhD student in mathematical physics at Ludwig Maximilians University in Munich, Germany.
e-mail: schaal@math.lmu.de

A RESEARCH
CAREER IS
POSSIBLE
IF YOU WANT IT
AND HAVE SUPPORT.
BUT IT COULD BE
A BIT EASIER.

SEVEN DAYS

The news in brief

BUSINESS

Publisher listing

Academic publisher Springer Nature has announced its intention to raise €1.2 billion (US\$1.5 billion) from selling new shares in an initial public offering on the Frankfurt stock exchange. Springer Nature, which publishes this journal and more than 3,000 others, was formed in 2015 in a merger between the London-based company Macmillan Science and Education and Berlin-based publisher Springer Science+Business Media. In 2017, Springer Nature posted revenues of €1.64 billion. Private German company Holtzbrinck Publishing Group, which owns 53% of Springer Nature, is planning to hold on to its stake. A large portion of the listing proceeds will be used to cut the company's debt, said a company spokesperson. (Nature's news team is editorially independent of its publisher.)

UNIVERSITIES

UK strikes off

A strike over pensions changes that would have affected tens of thousands of UK academics was suspended on 13 April, after union members voted to accept a deal from their employers. Staff at 65 universities walked out for a total of 14 days in February and March over changes that would have seen their pension income go from having a guaranteed element to being entirely dependent on investment return, leaving them worse off in retirement. Universities UK, which represents the employers, said the changes were needed to address a deficit in the pension fund. University and College Union (UCU) members last month rejected an initial offer to resolve the dispute. But nearly two-thirds

of voting members approved the second proposal, which commits to maintaining a guaranteed pension element. The plan would also create a joint panel to re-examine the pension scheme's valuation, which many academics say was inaccurate. The board of the pension scheme and the UK Pensions Regulator must now approve the proposal.

HEALTH

Flesh-eating illness

Australian scientists are calling for urgent scientific efforts to understand a worsening outbreak of Buruli ulcer, an infectious disease caused by a flesh-eating bacterium. In a 16 April letter in the *Medical Journal of Australia*, researchers report 236 new cases in the state of

Victoria from 1 January to 11 November 2017, compared with 156 from the same period in 2016 (D. P. O'Brien *et al. Med. J. Aust.* **208**, 287–289; 2018). Roughly 2,000 cases of the disease, which is caused by *Mycobacterium ulcerans*, are reported worldwide each year, mostly in West and Central Africa. The letter's four authors call for increased funding to investigate the environmental factors that assist the bacterium's growth, how it spreads to humans and why it is becoming more prevalent in Victoria. Mosquitoes, other insects and possums are considered possible vectors.

AI diagnostic tool

On 11 April, the US Food and Drug Administration gave the green light to a medical device

that uses artificial intelligence (AI) to detect diabetic retinopathy — a leading cause of blindness among people with diabetes in the United States. The agency says the device is the first AI tool it has approved that screens for disease without the need for a clinician to analyse the results. This makes the device suitable for use by health workers who are not eye-care specialists. It was developed by IDx, an AI company in Coralville, Iowa, and a clinical study found that it correctly identified people with diabetic retinopathy 87% of the time.

ENVIRONMENT

Polish logging

The European Court of Justice ruled on 17 April that Poland broke European Union wildlife



MARVIN JOSEPH/THE WASHINGTON POST/GETTY

March for Science back for round two

Supporters of science around the world took to the streets on 14 April for the second annual March for Science. Although more than 1 million people attended the 2017 events, the turnout this year was significantly smaller at many sites. Organizers say they held more than 250 marches, festivals and other events

this year, compared with the more than 600 demonstrations that took place last year. Among the cities holding demonstrations were Washington DC (pictured), Mumbai, Mexico City and Frankfurt — with many marchers turning out to call for increased research funding and to defend scientific values.

JOEL KOWSKY/NASA

laws by allowing increased logging in the country's ancient Białowieża forest. Polish authorities had in 2016 tripled logging limits to fight a pest outbreak in the forest, home to the largest population of European bison and a variety of rare birds and insects. Environmental groups filed a complaint with the European Commission, which referred the case to the EU's highest court. In July last year, the court ordered a preliminary ban on tree felling in the forest; despite fierce protests by campaigners, logging continued throughout the summer and autumn. The government said in December that it would comply with any ruling. After a government reshuffle, Poland's new environment minister suspended most logging in January, pending the court's final decision.

Pollution plan

More than 170 countries have agreed on a plan aimed at reducing greenhouse-gas emissions from the shipping industry, the United Nations International Maritime Organization (IMO) said on 13 April. The agreement calls for at least a 50% reduction in carbon emissions by 2050 compared to 2008 levels, with a long-term goal of phasing such pollution out completely. The IMO plans to flesh out and finalize a regulatory framework by 2023.

Emissions-reduction strategies under consideration include strengthening and extending energy-efficiency regulations that the IMO adopted in 2011, mandating cleaner fuels or new engine technologies, or imposing lower speed limits on ships in international waters to reduce fuel use.

TED awards

Efforts to explore the ocean's 'twilight zone' and launch a methane-monitoring satellite have won support from the Audacious Project, a philanthropy programme run by the non-profit group TED in New York City. On 11 April, the group announced that it has given US\$35 million to the Woods Hole Oceanographic Institution in Massachusetts to develop advanced robotic vehicles that can explore the area 200–1,000 metres below the ocean surface — an area where sunlight fades but life thrives. The Environmental Defense Fund, an advocacy group in New York City, will receive an unspecified amount in the tens of millions of dollars to develop a satellite that can monitor methane emissions from oil and gas fields. See page 283 for more.

PEOPLE

NASA science chief

Planetary scientist Jim Green (pictured) is NASA's new chief scientist, the agency



announced on 10 April. Green (pictured) has run NASA's planetary-sciences division since 2006, overseeing projects including three Mars rovers and the New Horizons missions to Pluto. NASA has not had a permanent chief scientist since geologist Ellen Stofan left in 2016. Stofan will become director of the Smithsonian Institution's National Air and Space Museum in Washington DC on 30 April.

Protein-fraud ban

An academic behind one of the most extensive cases of fraud uncovered in protein crystallography has received a ten-year government funding ban in the United States. On 10 April, the US Office of Research Integrity (ORI) said that H. M. Krishna Murthy, formerly an associate professor at the University of Alabama at Birmingham, had "falsified and/or fabricated" data in 12 entries in a protein-structure

database and in 9 research articles, including one published in *Nature* in 2006. The fraudulent research was supported by six grants from the US National Institutes of Health. In 2009, the University of Alabama at Birmingham reported the results of a two-year investigation into the case, and the ORI agreed with the findings. Murthy disputed the ORI's conclusions but a judge has now rejected his appeal. The judge's decision document details Murthy's defences, which include that he had made honest mistakes and that similar errors are common in research.

EVENTS

Spanish petition

Leading Spanish scientific organizations delivered a petition signed by more than 277,000 people to the national parliament in Madrid on 11 April, calling on the government to stop the "progressive abandonment of science in Spain". See page 285 for more.

Spy-poison probe

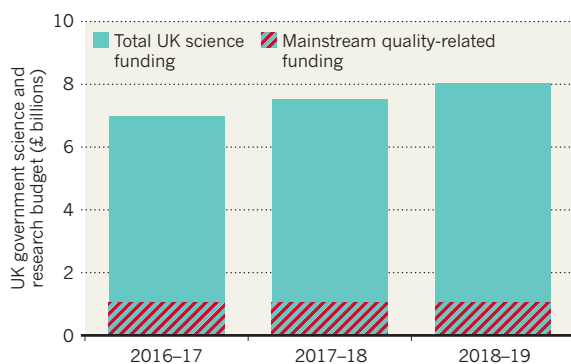
The international chemical-weapons watchdog has agreed with UK findings about the identity of the poison used in an attempted assassination of a former Russian spy in Britain. The Organisation for the Prohibition of Chemical Weapons (OPCW) carried out independent tests of samples from the town of Salisbury, where Sergei Skripal and his daughter Yulia were found poisoned in March. The OPCW has for now shared the name and the structure of the substance only with states party to the 1997 convention that bans the production or use of chemical weapons. Scientists with Britain's national defence laboratory at Porton Down say the compound belongs to a class of nerve agents called Novichoks. See page 285 for more.

TREND WATCH

The main pot of research funding given to English universities according to a nationwide quality evaluation known as the Research Excellence Framework (REF) will remain flat in 2018–19, although UK research funding more widely will rise. English universities will receive a total of £1.05 billion (US\$1.50 billion) in 'quality-related' funding to spend as they wish, the same as for the previous 2 years. But thanks to a funding boost announced in 2016, total cash for UK research is set to rise by 15%.

UK SCIENCE FUNDING

UK government research funding is set to increase by 15% compared with 2016–17 levels, but mainstream quality-related funding for English universities, determined by a research assessment, will remain flat.



SOURCE: RESEARCH ENGLAND

NATURE.COM

For daily news updates see:

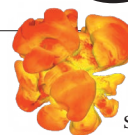
www.nature.com/news

NEWS IN FOCUS

INFECTIOUS DISEASE Deadly tick-borne virus spreads in East Asia **p.282**

ENVIRONMENT African nations gear up to collect air-pollution data **p.284**

CHEMICAL WEAPONS Attacks give impetus to boosting forensic capacity **p.285**



ASTROPHYSICS How to blow up a star: the science of supernova simulation **p.287**

RICHARD VEYERS/XL CATLIN SS/THE OCEAN AGENCY



Heat stress and bleaching — the loss of symbiotic algae — killed many corals in Australia's Great Barrier Reef after the 2016 crisis.

ECOLOGY

Great Barrier Reef saw huge losses from 2016 heatwave

One-third of the system's reefs were transformed after bleaching triggered by warmth.

BY QUIRIN SCHIERMEIER

Extrême heat in 2016 damaged Australia's Great Barrier Reef much more substantially than initial surveys indicated, according to ongoing studies that have tracked the health of the coral treasure. The heatwave caused massive bleaching of the corals that captured worldwide attention.

In a paper published on 18 April in *Nature*, researchers report¹ that severe bleaching on an unprecedented scale triggered mass death

of corals. This drastically changed the species composition of almost one-third of the 3,863 individual reefs that comprise the Great Barrier Reef.

The world's largest coral reef is unlikely to recover soon. The damage is a harbinger of what a warmer future might hold for a wealth of tropical reef ecosystems, says lead study author Terry Hughes, director of the coral-reef centre at James Cook University in Townsville, Australia. "If we fail to curb climate change, and global temperatures rise far above 2°C

[above the pre-industrial level], we will lose the benefits they provide to hundreds of millions of people."

FATAL LOSS

Hughes and his team of ecologists closely examined the 2,300-kilometre-long Great Barrier Reef after the 2016 heatwave. Extensive aerial surveys revealed widespread coral bleaching between March and April 2016. This phenomenon occurs when excessive heat kills or expels algae called ▶

► zooxanthellae, which have a symbiotic relationship with reef-building corals. The algae provide the corals with energy and nutrients from photosynthesis; without them, the corals often die.

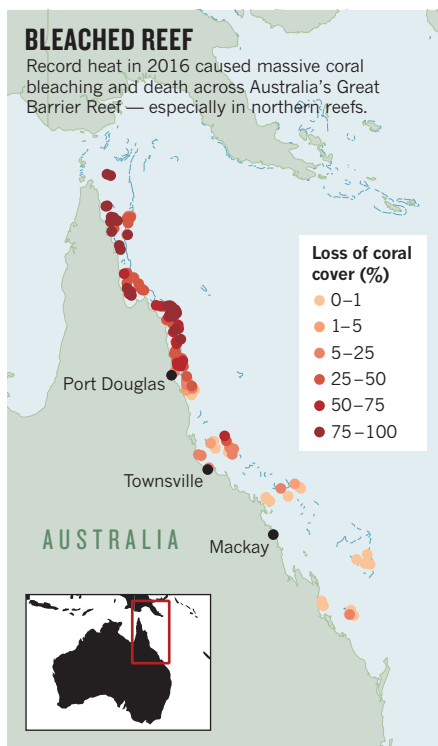
But to gauge the full extent of heat damage, Hughes's team conducted more-comprehensive underwater surveys of coral mortality, both at the peak of the observed bleaching in March and April, and again eight months later.

Many corals — especially those in the northern third of the reef — died immediately from heat stress. Others were killed more slowly, after their algal partners were expelled. The composition of coral assemblages on hundreds of individual reefs changed radically within just a few months of the heatwave. On severely bleached reefs, fast-growing coral species — which have complex shapes that provide important habitats — were replaced by slower-growing groups that shelter a less-diverse community.

"The study paints a bleak picture of the sheer extent of coral loss on the Great Barrier Reef," says Nick Graham, a marine ecologist at Lancaster University, UK. Approximately one-third of the world's coral reefs were affected by bleaching in 2016. On the Great Barrier Reef, less than 10% of reefs escaped with no bleaching (see 'Bleached reef'), compared with more than 40% in previous bleaching events studied.

"It is now critical to understand how governance and local management can maximize recovery between recurrent heatwaves," Graham says.

Tim McClanahan, a conservation zoologist



at the Wildlife Conservation Society in Mombasa, Kenya, says the study's findings might not predict how other reefs will cope with a warmer world. Responses might depend on the corals' life histories and local environmental conditions. "Global warming will result in more heat-stress events," he says, but "there is accumulating evidence that corals do acclimate".

Before the extreme 2016 incident, global coral bleaching had been observed just twice: in 1998 and 2002. Coral colonies can recover from such events, especially given that the species most susceptible to dying from heat stress are among the fastest-growing corals. But harmful warming events are occurring more frequently, and scientists think that full recovery is becoming increasingly difficult².

Researchers have also found that local protection of reefs and surrounding waters does little to make corals less sensitive to heat³. Rather, global changes such as ocean acidification might further increase environmental stress.

The fate of tropical coral reefs — including the iconic Great Barrier Reef — therefore depends on efforts to mitigate climate change, says Graham. "A future with coral reefs, their rich diversity and the livelihoods they provide to people is quite simple. It will only be possible if carbon emissions are rapidly reduced," he says.

But even if that happens, tomorrow's reefs might look different from today's, as the mix of species changes in favour of those that can best cope with inevitable climate change, says Hughes. "This transition is already under way, faster than many of us expected," he says. "The Great Barrier is shifting radically, a trend that will continue for the next century or more." ■

1. Hughes, T. P. *et al.* *Nature* <https://doi.org/10.1038/s41586-018-0041-2> (2018).
2. Hughes, T. P. *et al.* *Science* **359**, 80–83 (2018).
3. Hughes, T. P. *et al.* *Nature* **543**, 373–377 (2017).

SOURCE: REF. 1

INFECTIOUS DISEASE

East Asia braces for surge in deadly tick-borne virus

Rapid rise in number of infections concerns researchers.

BY DAVID CYRANOSKI

Infectious-disease experts in East Asia are preparing for this year's wave of a lethal tick-borne virus. The virus causes a disease called severe fever with thrombocytopenia syndrome (SFTS), which has affected a rapidly growing number of people since it emerged nearly a decade ago.

Scientists in the region say they are worried by the rising incidence of the disease, and by signs that the virus can spread more easily than previously thought. In March, Japan launched the first clinical trial of a drug to treat

the infection, and some researchers say that governments should devote more resources to raising awareness and studying the virus.

"It is our responsibility to come up with an effective treatment," says Masayuki Saijo, a virologist at the National Institute of Infectious Diseases in Tokyo, who helped to launch the trial.

Cases of SFTS were first reported in China in 2009 (X.-J. Yu *et al.* *N. Engl. J. Med.* **364**, 1523–1532; 2011). Researchers identified the virus responsible in blood samples from a cluster of people who shared a combination of symptoms, including high fever, gastrointestinal

problems, low white blood cell count and low platelet count (thrombocytopenia).

The virus killed 30% of those infected in China that year. It was even more lethal when the first cases appeared in Japan and South Korea in 2013. More than one-third of those infected in Japan and nearly half of those infected in South Korea died that year.

And the number of cases in each country has risen sharply. In 2013, there were 36 reported cases in South Korea; by 2017, the number had jumped to 270. In 2010, China reported 71 cases; in 2016, there were around 2,600. Japan experienced a 50%



The tick *Haemaphysalis longicornis* transmits an emerging, potentially fatal virus to people.

increase between 2016 and 2017.

All three countries have implemented measures aimed at educating local physicians and citizens in endemic areas about the risks of tick bites. Those infected now fare much better. In China, only around 3% of people infected died in 2016, and in Japan the number fell to 8%. In South Korea, the figure dropped from 47% in 2013 to 20% in 2017. Scientists credit the reduced fatality to earlier recognition and better general treatment — although no cure exists — and to the likelihood that wider surveillance has led physicians to recognize mild as well as severe cases.

The SFTS virus is not expected to evolve into a rapidly transmitted disease like Ebola. And infections are generally limited to people, such as farmers or hunters, who come into contact with the animals that carry *Haemaphysalis longicornis*, the tick that harbours the virus.

But many say that the virus's toll and potential threat have been under-appreciated. Those infected have a better prognosis, but the virus still kills a higher percentage than any other infectious disease in South Korea, says Keun-Hwa Lee, a microbiologist at Jeju National University in South Korea. And the higher number of infections means that the disease claims more than 100 lives globally each year.

Many animals, including goats, cattle, sheep and deer, expose humans to the ticks, and are often infected without showing symptoms. Current control efforts that focus on known endemic areas could fail, says Bao Chang-jun, a biostatistician at Jiangsu Provincial Center for Disease Control and Prevention in Nanjing. The course of the

epidemic “may change with human activities and climate change”, says Bao. “It’s necessary to conduct research on potential risk areas.”

Two reports from Japanese health officials last year caused particular alarm. One stated that a woman had been infected through a cat bite, and the other that a man had been infected by his dog. “To the warnings of previous years, we have to add the risk of touching sick domestic animals,” says Kazunori Oishi, director of the Infectious Disease Surveillance Center in Tokyo.

CLINICAL TRIAL

Last month, Japan began a clinical trial of an influenza drug, favipiravir, that was used to treat Ebola during the 2014 outbreak in West Africa. The drug is effective on viruses with a certain molecular structure that Ebola and SFTS share, says Saijo.

Although the number of cases has risen sharply, scientists can’t say whether the increase is due to heightened surveillance and awareness, a real growth in the number of ticks and the animals that carry them, or an increase in risk as humans encroach on areas where the disease is endemic. Shigeru Morikawa, director of the department of veterinary science at Japan’s National Institute of Infectious Diseases, says that some researchers suspect the number of ticks has increased because fewer people hunt wild animals in Japan now, and this has allowed deer and boar populations to surge.

Researchers say they have many questions about the virus and how it spreads, but they suspect that the chances to study the disease will go up soon, as warm weather returns and people flock to the outdoors, where they can come into contact with the ticks. “There will be more cases,” says Hideki Hasegawa, a pathologist at the National Institute of Infectious Diseases. “The season is just beginning.” ■

CLIMATE CHANGE

Big prize for methane probe

US green group wins millions to launch satellite.

BY JEFF TOLLEFSON

An environmental group in the United States has been awarded tens of millions of dollars to develop a satellite to help track emissions of the greenhouse gas methane from oil and gas facilities around the world.

If the Environmental Defense Fund (EDF) succeeds in launching its probe, it could be the first environmental organization to send its own satellite into space. Its work is being funded through the Audacious Project, a joint effort of the non-profit group TED and philanthropic organizations such as the Bill & Melinda Gates Foundation.

The EDF, based in New York City, aims to launch ‘MethaneSAT’ as early as 2020, with the help of scientific partners at Harvard University and the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. The group says that the probe will make the most-precise measurements yet of methane from space, and its data will be freely available.

The oil and gas industry emits around 76 million tonnes of methane globally each year, according to the International Energy Agency in Paris. That’s enough to power about 285 million US homes. The EDF’s goal is to monitor emissions from roughly 50 sites that account for around 80% of the world’s oil and gas production. But the satellite could also be used to estimate emissions from landfills and agriculture. “We need good solid data so that we really can support global action on climate change, and we’ve got to do it fast,” says Steven Hamburg, the EDF’s chief scientist.

The most-detailed measurements of atmospheric methane concentrations available come from the European Space Agency’s Sentinel-5P spacecraft, which was launched in October 2017. It provides global coverage at a resolution of nearly 50 square kilometres.

The EDF team is designing MethaneSAT to provide measurements at a resolution of 1 square kilometre, with global coverage at least once a week. That information can then be plugged into atmospheric models to calculate cumulative emissions across larger areas, says Steve Wofsy, an atmospheric scientist at Harvard who is working on the project.

“EDF has a very good team, and I have no doubt that it can be done,” says Charles Elachi, who formerly headed NASA’s Jet Propulsion Laboratory in Pasadena, California. “The challenge is how much it’s going to cost.” ■



Cooking stoves that burn biomass are a significant source of air pollution in sub-Saharan Africa.

PUBLIC HEALTH

Africa study seeks to fill pollution data gap

Low-income countries in sub-Saharan Africa are nearly unrepresented in the research on air quality and health.

BY NICOLE WETSMAN

At ten elementary schools on the outskirts of Kampala, the capital city of Uganda, newly installed air-quality monitors are quietly collecting data on the amount of particulate matter in the atmosphere. The schools are part of a project launched in February to study how air pollution affects children's health, in an effort to address a major public-health gap in sub-Saharan Africa.

Globally, air pollution causes more deaths than any other environmental hazard (P. J. Landrigan *et al. Lancet* **391**, 462–512; 2018). But there are few data on its health effects in sub-Saharan Africa. And it's hard to draw any lessons from similar studies in Europe or North America, because much of the air pollution in sub-Saharan Africa comes from a different source — indoor stoves that burn biomass such as charcoal and firewood. The resulting emissions of particulate matter, carbon monoxide and sooty 'black carbon' "can be hazardous indoors and can also go

outside to mix with other sources of pollution", says Eric Coker, who studies global-health equity in Uganda at the University of California, Berkeley.

Lack of access to health care, low nutrition and the high prevalence of infectious diseases such as tuberculosis and HIV leave the region's population susceptible to the effects of environmental pollutants, Coker adds.

The paucity of data was clear to Kiros Berhane, a biostatistician and a principal investigator with the Eastern African GEO-

"This was the place we could make the biggest contribution."

Health hub, the group running the child-health study in Kampala. The hub, which began conducting research in 2016, is a collaboration between the University of Southern California in Los Angeles and Addis Ababa University in Ethiopia. It chose to focus on air pollution after examining gaps in public-health research in the region. "This was the place we could make the biggest contribution," he says.

South Africa is the only country in sub-Saharan Africa with an air-quality monitoring programme, says Coker. Yet the sparse data available for the rest of the region suggest that in some areas, average levels of one type of pollution — particulate matter — are an order of magnitude higher than those in North American and European cities, Coker and a colleague reported last month (E. Coker and S. Kizito *Int. J. Environ. Res. Public Health* **15**, 427; 2018).

ACTION ACROSS AFRICA

The Eastern Africa GEOHealth hub aims to begin filling the sub-Saharan Africa data gap. The project is one of seven GEOHealth programmes centred in low-income countries across the world, and is funded in part by the US National Institutes of Health and Canada's International Development Research Centre.

The programme's child-health study stationed its air-quality monitors at ten schools outside Addis Ababa for a year, before moving them to Uganda in February and March. The devices will stay there for about a year, measuring levels of fine particulate matter.

Researchers are also tracking the lung function of children in the schools, using questionnaires and breathing tests. Once they finish collecting data in Uganda, they'll move the air monitors to schools in Kenya and Rwanda. Ultimately, the study plans to gather data from 40 sites across the 4 countries, and track thousands of schoolchildren, Berhane says. It's modelled on similar research conducted in southern California. "The idea is to see if lung function is actually associated with high versus low particulate matter," he says.

The GEOHealth hub is also installing air-quality monitors in the capital cities of each of the four countries included in the school study. Project researchers plan to compare pollution levels to morbidity and mortality in each city's major hospitals.

Air pollution hasn't been a priority for governments in eastern Africa until the past few years, Berhane says. With only limited resources available, governments have given more attention to concerns such as infectious disease and food insecurity. But attitudes have started to shift, and there is now more recognition of the damage that environmental exposures can cause.

The GEOHealth hub involves local stakeholders and government officials, and invites representatives to its training sessions and meetings. "They've been part of the process from the get-go," Berhane says. "It's translated into increased interest in the issue."

When the air-quality monitors in the children's health study were about to be moved from Ethiopia to Uganda, Berhane says, officials from Addis Ababa indicated that they were interested in replacing the monitors, and in continuing to track air quality. "I'm very optimistic that the work will continue," he says. ■

SCIENCE FUNDING

Huge petition in Spain decries 'abandonment' of research

Scientists deliver 277,000-signature campaign to parliament in protest over poor funding.

BY MICHELE CATANZARO

Spain's leading scientific organizations delivered a petition signed by more than 277,000 people to the national parliament in Madrid on 11 April, calling on the government to stop the "progressive abandonment of science in Spain" caused by budget cuts. The petition is the largest ever on a science-policy subject in Spain.

Its key promoters urge the government to, by 2020, return investment in science to almost €10 billion (US\$12.3 billion) — a level last seen in 2009.

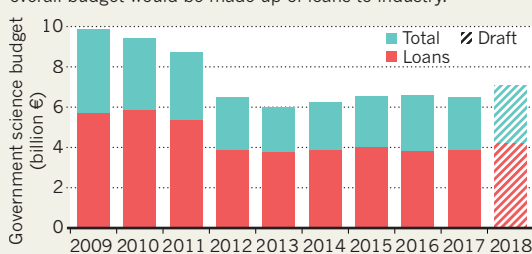
Between 2009, when the global financial crisis hit, and 2013, the country's science budget plunged by 39%, to about €5.9 billion (see 'Spain's science woes'). Since then, research funding has increased little.

A group of researchers posted the petition online in February. Its backers include the Federation of Young Researchers and the two largest Spanish workers' unions.

The petition's delivery to parliament came a week after Spanish scientists received a glimmer of good news. On 3 April, the government

SPAIN'S SCIENCE WOES

On 3 April, the Spanish government announced plans to boost its science budget by 8.3% in 2018, but about 60% of the overall budget would be made up of loans to industry.



submitted a draft budget for 2018, which included an 8.3% boost for science funding over 2017's allocation. If approved by parliament, the raise would be the biggest in a decade, and would take the budget to around €7 billion.

"We welcome the increase. It's an essential step, but just the first one," says Nazario Martín, a chemist at the Complutense University of Madrid and president of the Confederation of Spanish Scientific Societies (COSCE).

Some scientists still see problems. For the past several years, much of the science budget

has been made up of loans. And many scientists have accused the Spanish government of using this to inflate science funding. The loan money is usually offered to companies for applied-research projects but must be returned, although under favourable interest conditions. However, a substantial part of the loans usually goes unrequested and remains unspent. An analysis of the budget released by COSCE last week shows loans would make up about 60% of the overall 2018 science budget

There's no guarantee that the budget will be approved in its current form, and the governing People's Party does not hold an absolute majority in the parliament. But Alicia Durán, a physicist at the Spanish National Research Council (CSIC) who was one of about 65 scientists who delivered the petition, is hopeful that it will effect some change. Members of parliament across the political spectrum have said that they would be prepared to sign an amendment to the budget protecting aspects of science they all agree on — such as improving funding — from negotiations about other parts of the budget, says Durán. ■

SOURCE: ICONO-FECYT/COSCE

WEAPONS

Chemical attacks highlight need for better forensics

International watchdog probes assassination attempt in Britain and suspected Syria attack.

BY DECLAN BUTLER

As investigations continue into the attempted assassination of a former Russian double agent and his daughter in Britain, findings released this week have renewed focus on the class of nerve agents allegedly used. And experts say that the UK event and a suspected chemical-weapons attack in Syria provide fresh impetus for international efforts to beef up forensic capabilities.

On 12 April, the Organisation for the

Prohibition of Chemical Weapons (OPCW) confirmed that its independent tests of environmental and biological samples related to the assassination attempt identified the same poison as did investigations by forensic scientists at Britain's national Defence Science and Technology Laboratory at Porton Down. The attack happened in the nearby city of Salisbury on 4 March. The OPCW, based in The Hague in the Netherlands, is responsible for enforcing the 1997 Chemical Weapons Convention, which bans the production and use of such arms.

The organization did not name the chemical agent publicly, but will share its identity and structure with states party to the convention, in a classified report. More details might emerge at a special meeting of the OPCW's executive council to discuss the report, scheduled for 18 April. The UK government has said that the compound belongs to a class of nerve agents known as Novichoks.

The watchdog also agreed that the toxic chemical was very pure. That points to it having being made by "a highly proficient team and in a well-refined process", says ▶

► Alastair Hay, an environmental toxicologist at the University of Leeds, UK. The OPCW did not say where the substance might have been made; the UK government has alleged that the Russian state was directly behind the attack, but critics say that this is a politically motivated claim and that there is no forensic evidence to back it up.

CHEMICAL DETECTIVE WORK

Experts say further investigations could provide more clues. Forensic inquiries into chemical attacks typically involve standard tools such as gas and liquid chromatography, which are used to separate a substance into its components. Researchers then study those compounds with analytical techniques such as high-resolution mass spectrometry, nuclear magnetic resonance spectroscopy, isotope-ratio mass spectrometry and inductively coupled plasma mass spectrometry.

Forensic methods can build up chemical signatures of the components of a sample to give investigators leads about how it was made, says Brad Hart, director of the Lawrence Livermore National Laboratory's Forensic Science Center in Livermore, California. The varying ratios of stable isotopes of component elements, for example, can provide information about where the starting materials came from, he says.

Other sample components can offer clues

about the methods of synthesis, potential starting materials and the sophistication of manufacture, he adds. "Anything detected in the sample that is not the primary prod-

Forensic methods can build up signatures of a sample to give leads about how it was made.

uct is of interest as a potential signature," says Hart. "These typically include unreacted starting materials, products of side reactions, breakdown or decomposition products of the primary product or other signatures."

But it's not yet possible to definitively identify the geographical or institutional source of a chemical weapon using chemical forensics alone, he says.

ALLEGED ATTACK

The OPCW report came just days after an alleged chemical-weapons attack on the city of Douma in Syria, on 7 April. An OPCW fact-finding mission has gone to Syria to investigate the incident. The team will interview witnesses, and collect samples and evidence such as autopsy reports and photographs. Experts say that such attacks underscore the need to increase international chemical-forensic capacity for investigation, and intensify research in the field.

The OPCW is already taking steps in this direction. From 12 to 14 February, it held the first meeting of its science board's newly created temporary working group — made up of leading scientists and experts from national defence and other labs — charged with carrying out an in-depth review of the state of the art of chemical forensics.

And in April last year, international researchers, treaty experts, law-enforcement agencies and industrialists formed the Chemical Forensics International Technical Working Group, an ad hoc group aiming to identify research gaps and other factors that hinder investigators who use forensics to track down the source of chemical weapons.

A first glimpse of the panel's plans came at the OPCW's February science meeting. Carlos Fraga, a chemical-weapons specialist at the Pacific Northwest National Laboratory in Richland, Washington, and a driving force behind the international technical working group, says the team is proposing to develop a database of signatures of chemical weapons and their precursors. During the destruction of the world's chemical-weapons stockpiles, researchers gathered vast amounts of analytical data that could be added to the database, along with unpublished data collected during the OPCW's routine inspections of chemical plants, and by OPCW-designated labs. ■

LAST SECONDS OF STARDOM

BY ELIZABETH GIBNEY

Three-dimensional simulations are closing in on a 50-year-old problem: **what makes massive stars explode when they die?**

After spending three months trying to blow up a star, Hans-Thomas Janka and his team finally saw what they had been waiting for. Like the world's most patient pyromaniacs, they watched their massive stellar simulation — rendered in painstaking detail — inch closer to detonation. Each day, their supercomputer ticked through just 5 milliseconds of the star's life.

But perseverance has its rewards. In the team's previous attempts to make a realistic simulation, the stellar fireworks always petered out. This time, in 2015, Janka watched as the shock wave needed to drive the explosion continued to grow; the mock star was going supernova¹. “That was the moment we recognized that, OK, now we are at the point we longed to be at for two decades,” says Janka, a theoretical astrophysicist at the Max Planck Institute for Astrophysics in Garching, Germany. “We were on the path to clarifying the explosion mechanism of these massive stars.”

For more than half a century, physicists have suspected that the heat produced by elusive particles called neutrinos, created in the core of a star, could generate a blast that radiates more energy in a single second than the Sun will in its lifetime. But they have had trouble proving that hypothesis. The detonation process is so complex — incorporating general relativity, fluid dynamics, nuclear and other physics — that computers

An artist's impression of supernova 1987A, showing its asymmetric ejection of material.

have struggled to mimic the mechanism *in silico*. And that poses a problem. “If you can't reproduce it,” Janka says, “that means you don't understand it.”

Now, improvements in raw computing power, along with efforts to capture the stellar physics in acute detail, have enabled substantial progress. Janka's simulation marked the first time that physicists had been able to get a realistic 3D model of the most common type of supernova to explode. Just months later, a competing group based at Oak Ridge National Laboratory in Tennessee repeated the feat with a heavier, more complex star². The field is now buzzing, with more than half a dozen teams currently working on exploding stars in 3D. Many researchers are confident that they are closing in on identifying the ingredients that are crucial to generating such blasts.

The effort faces challenges. Three-dimensional models are still in their infancy and vary widely — and simulated stars sometimes still fail to blow. Time is also of the essence. Stellar explosions beyond the Milky Way are a common sight, but astronomers want to see one up close, in our own backyard. One or two are expected to happen every century, and the next one could occur at any time. When it does, astronomers will be equipped to see more than just the light emanating from the outer layers of the explosion. They will be able to use state-of-the-art detectors to pick up gravitational waves and neutrinos emanating from the centre of the blast. Not only can predictions from simulations help astronomers to tailor their instruments to best capture the explosion, but they will also be essential for making sense of the data. “My goal is to have the models sufficiently sophisticated so that when a Galactic supernova goes off, we're ready for it,” says Anthony Mezzacappa, who leads the Oak Ridge team.

BEHIND THE SHOCK

When a star between around 8 and 40 times the mass of the Sun comes to the end of its life, it tends to go out with a bang, releasing more energy than one trillion trillion nuclear warheads. These “core collapse” explosions

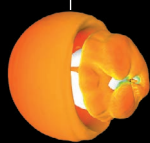
LUIS CALÇADA / EUROPEAN SOUTHERN OBSERVATORY

EXPLODING A VIRTUAL STAR

When a massive star dies, a neutron star can form at the centre of its iron core. Infalling iron hits this ultra-dense orb and rebounds, creating a shock wave. In simulations such as this one, of a 20-solar-mass star, the crucial moments occur right after the bounce.

Tiny perturbations in the flow of matter through the shock wave amplify into violent sloshing motions around the neutron star (white)

Heating from neutrinos produced in the neutron star also causes bubbles of convection in the infalling matter, which build pressure behind the shock wave



170 ms
(after rebound)



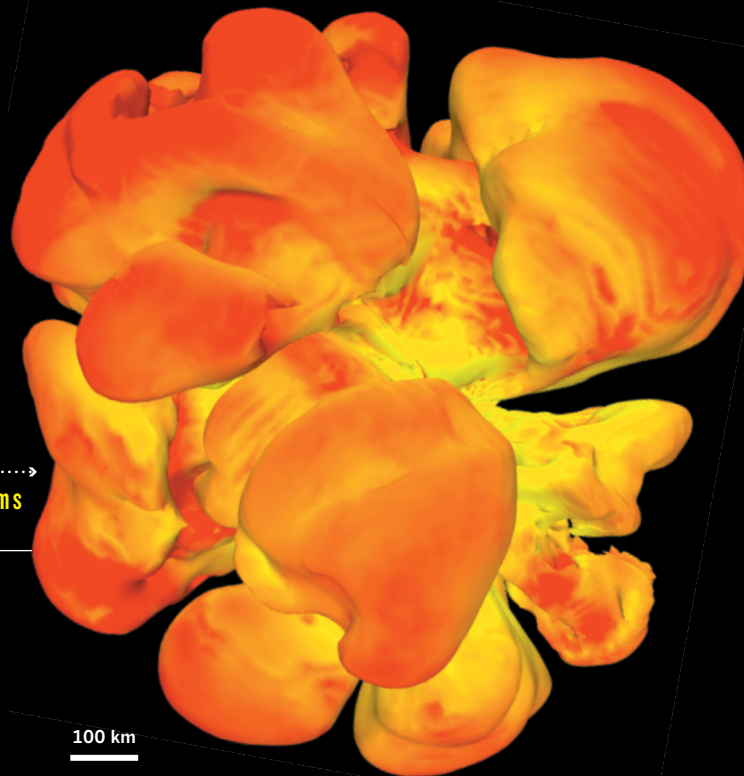
280 ms

527 ms

RELATIVE VELOCITY OF MATTER



Heat from neutrinos and the pressure of these motions eventually drive the shock wave rapidly out into layers beyond the core — the star is exploding



make up around two-thirds of all supernovae. (The other sort, known as type Ia, involves a fusion-driven explosion of a white dwarf.)

Interest in core-collapse supernovae began in the late 1950s, when scientists first theorized that a range of chemical elements — including most of those crucial to life — are forged in stars. Some of the heaviest elements, they thought, would arise in the high-energy, rapidly evolving furnace of a supernova³. The explosion would then spew them out, seeding space with the ingredients for stars and planetary systems.

Astrophysicists think that, before they explode, these stars run low on gas — namely hydrogen. With less to fuse, an old star no longer generates as much radiation, and its core contracts under gravity. Lighter elements progressively fuse into heavier ones, but stop short at iron. Ultimately, unable to resist gravity, the centre of the iron core collapses in a fraction of a second into the densest type of matter known: a neutron star.

It is commonly thought that infalling matter then hits the newly formed neutron star and rebounds, creating a shock wave that ripples out from the centre. But the rebound alone is too weak to both reverse the collapse of material and send the outer layers of the star flying. Without some extra source of energy, it stalls on its way out. This shortfall, Janka says, “has been puzzling us for more than 50 years”.

Solving the puzzle — and understanding the dynamics of the particle soup at the star’s heart — is crucial to working out how atomic elements form and in what abundance, Janka says. It could also help to determine when a star might collapse into something even more exotic, such as a black hole. “These questions are not understood without deciphering the explosion physics,” says Janka. And there’s another reason that modellers tackle the question, adds Sean Couch, a computational astrophysicist at Michigan State University in East Lansing. “I think if you pushed lots to tell the truth, we just really like blowing things up,” he says.

But the question of what makes a star explode has stood for more than half a century because it is almost unfathomably hard — and computers have not been powerful enough to tackle the problem, says Maryam Modjaz, an astrophysicist at New York University. “It’s one of the most complex systems that we can model,” she says. Physics at seemingly every scale comes into play, from the bending of space-time to the particle physics of neutrinos and the behaviour of matter under extreme

pressure. Getting to the state of today’s simulations, and the still-tentative explanations of how core-collapse explosions happen, is a decades-long story of increasing complexity that began with something that looked little like a star: simple 1D models.

Although fairly crude, those models revealed the first vital ingredient of a core-collapse supernova: the neutrinos produced through particle interactions in the newly formed neutron star. Neutrinos, which are nearly massless, barely engage with other particles. But in 1966, theorists calculated that if even a tiny fraction of their energy was absorbed by the dense matter around the core, the heat would be enough to rekindle the shock wave and drive it out⁴. Evidence in favour of the idea might have been bolstered by a lucky break. In 1982, computational physicist James Wilson, then at Lawrence Livermore National Laboratory in Livermore, California, left a simulation to run overnight — some say accidentally. He returned to find that, after a delay, enough neutrinos had diffused out of the neutron star to heat matter behind the shock wave and drive it out of the star. Until then, physicists had not realized that a stalled wave could be revived. “If models were not run to such late times, we would not have seen it,” says Mezzacappa.

Neutrino heating became the field’s main focus of research, but the more detailed the simulations and the larger the mass of their starting star, the less often modellers saw explosions. Although neutrinos pushed the stars close to the brink, it became clear that they needed a helping hand.

FULL FIREWORKS

The first clue as to what might provide the boost came in 1987, when astronomers observed a supernova in a nearby galaxy — the Large Magellanic Cloud. At the time, 1D models necessarily assumed stars were perfect spheres, made up of concentric layers of fusing elements and containing dynamics that could be captured with just one coordinate: distance from the centre. But the intermingled way that supernova 1987A spewed out elements suggested that layers must mix, a dynamic process that would be impossible to describe in one dimension.

With the advent of much more powerful computers in the 1990s, modellers were able to capture this motion by progressing from 1D to 2D simulations. In two dimensions, neutrino heating acted like a stove flame

SIMULATION: T. MELSON (MAX PLANCK INSTITUTE FOR ASTROPHYSICS, MPA) VISUALIZATION: AARON DOERING (MPIA) AND ELENA VASSEROT (MAX PLANCK COMPTON & DATA FACILITY) SOURCE: T. MELSON ET AL. *ASTROPHYS. J.* **808**, L42 (2015)

under a pan of water, creating convection and turbulence that churned up fresh matter for the particles to heat, boosting the pressure behind the shock wave. And in 2003, Mezzacappa's team found that perturbations in the shock wave can grow rapidly into large sloshing motions and violent rotations — known as standing accretion shock instability (SASI). These motions charge the shock wave and help the star to explode.

Still, physicists worried that the compromises they made in rendering stars in two dimensions might artificially boost the chance of explosion. Indeed, when computing power made crude 3D models feasible in the early 2010s, the models were once again “reluctant to explode”, says Bernhard Müller, a computational astrophysicist at Monash University in Melbourne, Australia, who was part of Janka's team until 2014. It was not until the advent of faster supercomputers in 2012 that researchers began to be able to weave together general relativity and detailed nuclear and particle physics to get 3D stars to start to blow, in models that ran from scratch.

Reaching that milestone lends confidence to the assumption that neutrino heating, convection and SASI oscillations are behind the explosions, says Janka. Since 2015, teams around the world — including groups at California Institute of Technology (Caltech) in Pasadena, Princeton University in New Jersey, Michigan State University and Fukuoka University in Japan — have begun to work on 3D models. A substantial fraction of those simulations end in explosions (see ‘Exploding a virtual star’). The trend will need to continue across a range of stars of different masses and initial structures to prove that physicists understand the mechanism, but Müller is optimistic. “We seem to be converging towards a solution for this problem of shock revival,” he says.

Others are more sceptical. Shock waves emerge more easily in relatively small stars. When Janka's team attempted to explode a larger 3D star in 2015 — one that was 20 rather than 10 times the mass of the sun — they succeeded only because they pushed one interaction rate for neutrinos to the lowest level that the error bars from particle physics would allow. Today's simulations, which use more realistic initial conditions, still sit uncomfortably close to the tipping point between exploding and sputtering out, and no one is quite sure why. “In nature, these things explode robustly all the time,” says Couch. The models' reluctance to do so is “probably telling us either we're not doing it accurately enough with the physics we are including, or we're missing physics”.

A solution is to keep building richer models. But on today's supercomputers — which perform the equivalent of tens of thousands of state-of-the-art home computers running at once — this process still takes months, and modellers must necessarily make approximations and simplifications. Upgrades due in the next few years to supercomputers in the United States, Europe and Japan would cut the run time for a 3D explosion down to weeks. But even after that, computers would need to be made 100 times more powerful to churn through a 3D simulation that takes into account the full complement of physics, says Mezzacappa. Such computers could be another decade away, he says.

In the meantime, physicists are focusing on adjusting their models to see whether they can work out how the three main ingredients — neutrino heating, convection and SASI oscillations — interact, and whether any others might be missing. Some are exploring whether rotation and magnetic fields might help fuel the explosion. Others are basing models on more-realistic stars, with perturbations built in from the start. But comparing across simulations is difficult. Each group's models include not just different physics, but different shortcuts, resolution and pixel geometry — all of which can affect the result. And teams defend their choices fiercely. “I would go to conferences and people from different groups were almost fighting with each other, each saying ‘my code is better,’” says Modjaz. “There was no way to tell, because they wouldn't publish their codes or compare them in a regular fashion.”

Now groups are realizing that to make progress, they might need to find ways to make those comparisons, says Modjaz. A new generation of modellers, including Couch and Evan O'Connor at Stockholm University, have pioneered the publication of codes and encouraged others to do the same. Janka advocates creating a set of standardized test problems, with the same well-defined ingredients and initial conditions, to be used by the whole

field. “I think it will be a next very important step for the community, to enhance its credibility and the reliability of the results put out,” he says.

CORE QUESTIONS

The true test will be whether these explosions actually resemble the ones in nature. Models are now sophisticated enough — and computing capacity is great enough — to run simulations beyond the first fraction of a second after the shock wave forms to when the blast wave ultimately breaks through the surface of the star many hours later. The predictions of supernovae shape, energy and chemistry generated by such models can then be compared with a real star's exploding outer layers, as well as with the motion of the leftover core.

But studying light from the star's surface — as well as ghostly remains that linger for centuries — can give only limited information about the explosion. “It's like going to a dermatologist to ask about your heart,” says Couch. Neutrinos and gravitational waves, which pass through matter relatively unimpeded, could allow astronomers to see deep inside the star. In 1987, three neutrino detectors picked up 25 neutrinos emitted from supernova 1987A. In the decades since, subsequent detectors — such as IceCube at the South Pole and Super-Kamiokande in Japan — have been built that could be sensitive to tens of thousands of neutrinos emitted by a nearby supernova. When the neutrinos of such an explosion reach Earth, their energy, abundance and emission rate could reveal, for example, roughly how massive and how compact the neutron star is, as well as how much mass it continued to accrete after collapse. Any SASI wobble would cause neutrino emissions to rise and fall, and be visible as oscillations in the signal. “You could have a direct smoking gun for what's going on inside the supernova,” says Müller.

The value of detecting a supernova through its neutrinos is so great that upgrades to IceCube are usually done on only part of the detector at a time, so that it won't miss a once-in-a-lifetime event. The youngest supernova remnant found so far in our Galaxy is about 150 years old, but researchers say that it would be a statistical fallacy to think the next explosion is ‘overdue’. “No one can tell you when it will take place, so you have to be alert all the time,” says Janka.

If astronomers get lucky, the Laser Interferometer Gravitational-Wave Observatory (LIGO) in the United States and its sister observatory Virgo near Pisa, Italy, should also be able to observe the blast, although the signal is not expected to be as clear as those of the black-hole and neutron-star mergers found so far. Sarah Gossan, a physicist at Caltech and a member of the LIGO team, says that simulations will be needed to help find a faint signal among the noise and to decipher the information it contains. “We'll be able to inform our simulations from our observations, and vice versa,” Gossan says.

To prepare for such events, modellers such as Janka will need to simulate dozens of different 3D stars. In October, his team lit the fuse on a particularly complex model — a 19-solar-mass star, whose final minutes they had also modelled so that they could begin the collapse under conditions as messy and realistic as possible. They won't find out until at least July whether it will blow or not. But “by now”, he says, “we're pretty used to being patient”. ■

Elizabeth Gibney is a senior reporter for Nature based in London.

1. Melson, T., Janka, H.-T. & Marek, A. *Astrophys. J.* **801**, L24 (2015).
2. Lentz, E. J. *et al. Astrophys. J.* **807**, L31 (2015).
3. Burbidge, E. M., Burbidge, G. R., Fowler, W. A. & Hoyle, F. *Rev. Mod. Phys.* **29**, 547–650 (1957).
4. Colgate, S. A. & White, R. H. *Astrophys. J.* **143**, 626–681 (1966).



Chronotherapy — the specific timing of drug delivery — has shown promise in clinical trials. But that may not be enough to overcome the practical challenges.

TIME TRIALS

BY LYNNE PEEPLES

Carole Godain remembers a lot of the little details from the clinical trial she took part in nine years ago. There was the blue button she pushed to get her chemotherapy drugs, and the green light that came on to confirm that the medication was dripping into her veins. Then, of course, there was the hour — 10:00 p.m. without fail, for every treatment.

By all accounts, Godain's own time was running short. The first treatment for her colon cancer had failed, and her last body scan had revealed 27 tumours growing inside her liver. So the psychologist from Tours, France, jumped at the opportunity to take part in a trial at Paul Brousse hospital in Villejuif, which aimed to test whether delivering drugs at a specific time of day might make them more effective or reduce their toxic side effects. Ideally, it would accomplish both. "I was interested in increasing my chances of being cured," says Godain.

Today, at the age of 43, she is cancer-free. And Francis Lévi, the oncologist who treated Godain, says that

although such an amazing result is anomalous, emerging evidence should encourage more interest in the concept of chronotherapy — scheduling treatments so that they provide the most help and do the least harm.

More than four decades of studies describe how accounting for the body's cycle of daily rhythms — its circadian clock — can influence responses to medications and procedures for everything from asthma to epileptic seizures. Research suggests that the majority of today's best-selling drugs, including heartburn medications and treatments for erectile dysfunction, work better when taken at specific times of day. "When you give a medication, you always know the dose," says Lévi, who also now works at Warwick Medical School in Coventry, UK, where he leads a team associated with INSERM, the French national biomedical research agency. "We have found that the timing is sometimes more important than the dose."

Yet chronotherapy, sometimes called chronomedicine, remains at the fringes of clinical practice and drug-development programmes; the reasons for that are varied. Until about a decade ago, scientists could not explain the

ILLUSTRATION BY VIKTOR KOEN

molecular underpinnings for these circadian effects. And clinical data have been inconsistent — although a couple of Lévi's early trials, for example, showed clear benefits for people taking timed treatments, a later, larger trial produced more mixed results. Most patients haven't been as fortunate as Godain.

Axel Grothey, an oncologist at Mayo Clinic in Rochester, Minnesota says that the challenges facing chronotherapy are twofold: "You need more solid data. And you need to show it is feasible." The strategy can be impractical for cancer therapies, he says. Seats in his chemotherapy unit book up in much the same way as those for a movie. "The 4 p.m. showing could be oversold because we have too many patients who need to be started at that time," Grothey says.

Still, Lévi and others are optimistic. Chi Van Dang, the scientific director of the Ludwig Institute for Cancer Research, a global non-profit research organization, has noticed what he calls a "rebirth of interest" in chronotherapy, spurred by the rapidly advancing science of circadian rhythms and a handful of trials and technologies aimed at tailoring the approach to people's individual circadian clocks. These efforts could help to elucidate inconsistencies in clinical trials and make chronotherapy more practicable for doctors and patients alike, Lévi argues. Dang gave a keynote address at a chronotherapy workshop held by the US National Cancer Institute (NCI) last September. As the world's largest funder of cancer research, the NCI had put out a call a few months earlier for studies looking into how circadian processes influence disease progression and response to treatment. "I would argue that the evidence shows there is a benefit and we can't ignore it," says Dang. "We just need to be more clever on how to approach the challenges."

CLOCK WATCHERS

Chronotherapy enjoyed a publicity boost of sorts last year. Just a week after the NCI workshop, the Nobel Prize in Physiology or Medicine was awarded to a trio of scientists for elucidating the cellular mechanisms that control circadian rhythms. The circadian clock is a remarkable system. A central timekeeper in the hypothalamus orchestrates a network of peripheral clocks in nearly every organ and tissue of the body, turning on and off a bevy of genes, including some that encode the molecular targets for drugs and the enzymes that break drugs down. These clock genes are particularly important in cancer because they govern cell cycles, cell proliferation, cell death and DNA damage repair — all processes that can go haywire in cancer.

Some, but not all, cancers live by the clock as well, and researchers are trying to exploit their daily rhythms. When Joshua Rubin, a neuro-oncologist at Washington University School of Medicine in St. Louis, and his colleagues wanted to launch a chronotherapy clinical trial on a common and deadly form of brain tumour known as glioblastoma, they needed to check how the cancer behaved over time. So his team engineered cells from patient tumours to express luciferase — the protein that makes fireflies glow — every time core clock genes switched on. Then they watched. "It was so dynamic," says Rubin. "Lights go on, lights go off. Lights go on, lights go off." The team started treating the tumour cells with drugs at different times in the cells' cycle and found that they were most sensitive to an oral drug, temozolomide, near the daily peak in expression of the core clock gene *Bmal1* (ref. 1). If patients could be directed to take this pill — part of the standard glioblastoma treatment — at the time of peak *Bmal1* expression, the drug might be more effective, Rubin reasoned. His team is now testing that hypothesis in mouse models, and in more than

two dozen humans being treated at different times of day.

The trial is the first to apply chronotherapy in glioblastoma, and the only current trial in the United States that accounts for the circadian clock in cancer. A few previous US trials hinted that chronotherapy could be beneficial in treating ovarian², breast³ and non-small-cell lung⁴ cancers. Yet today, of the tens of thousands of ongoing clinical trials around the world, only a small fraction of 1% incorporate time-of-day considerations, according to a 2016 survey⁵.

The prospect nevertheless has some people excited, in part because of its simplicity. "If we can help people live longer and live better with fewer side effects, just by changing our scheduling, that would be tremendous," says Jeremy Rich, a neuro-oncologist at the University of California, San Diego. And the findings have intuitive appeal. Steroid levels, for example, naturally cycle with the circadian clock. In the late 1960s, scientists found that the synthetic corticosteroid methylprednisolone is safer for treating arthritis and asthma if taken in the morning rather than at other times of the day. This is because the feedback loop in the hypothalamus, which controls the release of cortisol, is least vulnerable to inhibition in the morning. These rhythms might affect responses to radiation treatment, too, says Eric Holland, a neurosurgeon at the Fred Hutchinson Cancer Research Center. Holland has shown that corticosteroids can reduce the effectiveness of radiation therapy in humans⁶ and that there are optimal times to administer radiation in mice⁷.

In one of the most cited cancer chronotherapy studies so far, Lévi and his team randomly assigned 186 people to either chronotherapy or standard treatment for colon cancer⁸. Slightly more than half of the people who, like Godain, had their chemotherapy infusion synchronized with their circadian rhythms responded to the treatment, compared with 29% of individuals on a standard schedule. And in a study published in January⁹, researchers found that for 298 patients randomly assigned to cardiac surgery in the afternoon, the subsequent risk of sustaining major heart damage was half what it was for 298 patients who underwent the same surgery in the morning. To avoid the possibility that the choice of surgeon was responsible for this difference, the study had the same doctors performing operations both in the morning and the afternoon.

The optimal time for various procedures seems to vary. Akhilesh Reddy, a physician-scientist at the Francis Crick Institute in London, suggests the cardiac surgery findings may translate to other surgeries — with prime times dependent on the peak expression levels of particular enzymes in respective tissues. For radiation treatment, Dang and other researchers have found mornings to be preferable to afternoons¹⁰. But as with the administration of chemotherapy, different types of tumours — and different people — may respond differently, says Dang. Lévi and others think that this might explain why many trials trying to reap the benefits of timed drug delivery have had more equivocal findings. The largest cancer chronotherapy trial so far — also led by Lévi — tested chronotherapy or conventional chemotherapy delivery in 564 people with metastatic colorectal cancer¹¹. Overall, it found that survival times were similar in both groups. But when results were broken down by sex, the risk of an earlier death dropped by 25% for men whereas it increased by 38% for women.

The reason behind those sex-related differences is not yet clear, although Lévi is starting to make some sense of them. His team presented findings in September 2017 suggesting that men best tolerate one type of cancer drug between four and seven hours earlier in the day than women do. Lévi also suggests that women experience more toxic effects, in

"WE HAVE FOUND THAT THE TIMING IS SOMETIMES MORE IMPORTANT THAN THE DOSE."

**“IT WAS SO
DYNAMIC.
LIGHTS GO
ON, LIGHTS
GO OFF.
LIGHTS GO
ON, LIGHTS
GO OFF.”**

general, than do men from cancer treatment.

Age is another factor that can affect an individual's rhythms. People's body clocks tend to shift in adolescence¹² — hence teens' preference for late nights and sleeping in — and back again as they age¹³. Overall, Lévi has found that about half of patients have similar circadian patterns. One-quarter have cycles that begin earlier, and the remaining quarter have ones that begin later — these two groups perhaps corresponding to the 'morning larks' and 'night owls' of the world. The bottom line is that there's no one-time-fits-all for chronotherapy.

CHALLENGING TIMES

Phyllis Zee, chief of sleep medicine at Northwestern University Feinberg School of Medicine in Chicago, says that chronotherapy has great potential, but that practical biomarkers are needed to help clinicians identify optimal times for treatment. “Those are the legs required for chronotherapy really to be translated,” she says. “It may not be ready for prime time.”

Lévi has been working on trying to track individual rhythms better. Before Godain started her home chronotherapy regimen, she strapped on a watch-like device that logged her daily rhythms, says Lévi. Godain had very regular sleep-wake cycles, which Lévi thinks probably contributed to her successful treatment. He and other researchers are now wielding even more sophisticated tools to discern circadian timing, including temperature sensors worn on the chest or ingested, blood samples and saliva tests. One research team at the University of Pennsylvania in Philadelphia is integrating data from wearable devices, smart phone apps and physiological samples in an effort to define each person's 'chronobiome' and pinpoint the best predictors for optimizing chronotherapy.

In some ways, chronotherapy could represent another arm in the effort to individualize treatments. “In the field of personalized medicine, adding this dimension of time could make a tremendous difference,” says Carla Finkelstein, a molecular biologist at Virginia Tech in Blacksburg. “We now have a really good molecular foundation. Hopefully, this is the beginning.”

Other practical challenges remain. Costs and convenience are at the heart of most scheduling decisions in a hospital. Bart Staels, a molecular pharmacologist at the University of Lille in France, and senior author of the cardiac surgery paper⁹, acknowledges that limiting heart surgeries to a certain time of day is not realistic. But doctors could identify patients at high risk of complications and prioritize them for afternoon surgery. Or perhaps clinicians could one day deliver a drug that artificially 'jet lags' a patient's heart into thinking a morning surgery is actually happening in the afternoon, says Staels.

Drug companies have been reluctant to take chronotherapy approaches for several reasons, says David Ray, an endocrinologist at the University of Manchester, UK. It can be difficult enough to get patients to take medication, regardless of time. Only about 50% of people with a chronic illness follow their treatment recommendations, according to the World Health Organization. What's more, regulators might insist that marketing a medication optimized for a specific time of day requires extra warnings about the risks of deviating from the schedule. That's not a good selling point for a liability-wary drug maker — nor is the price tag of running studies to show a time-based response. Twice as many study groups would be necessary to show that giving a drug at one time is better than giving it at another, says Ray. And for drugs already making money, companies

lack an incentive to go back and specify a time of day.

Ray and others also say that they are concerned by the trend among pharmaceutical companies to create once-a-day and other long-acting drug formulations. Doing so, they say, could have unintended consequences. Sustaining levels of a drug that targets the inflammatory molecule TNF- α , for example, could leave the immune system impaired throughout the day, says Ray. For conditions such as rheumatoid arthritis, he says, “you only really need to block TNF for a critical 4-to-5-hour window”.

John Hogenesch, a circadian biologist at the University of Pennsylvania, says that paying attention to the timing of treatments could eventually cut costs for companies. “I think what will change minds is showing people that when you take time of day into account, you can lower the noise and improve the signal between the controls and your clinical arm,” Hogenesch says. That could also mean rescuing some of the 90% of drug candidates that fail in early stages of development, says Ray.

As Rubin and his team look towards the next phase of their trial in the United States, they intend to measure participants' rhythms and give temozolomide accordingly, he says. Meanwhile, in Europe, researchers are using portable devices to track around-the-clock blood pressure in thousands of patients, to build on evidence that conventional hypertension medications are best dosed at night. A study¹⁴ published in February notes a 67% reduction in heart attacks, strokes and other major cardiovascular events among patients taking bedtime doses, compared with those on morning doses. Juan Crespo Sabarís, a physician affiliated with the University of Vigo, Spain, who has been involved with the hypertension work, noted that doctors in his region of Spain are now advising bedtime dosing as a simple, low-cost form of chronotherapy.

For champions of the approach, such as Lévi, the prospects for chronotherapy have never looked better. But given the mixed results from trials, and the practical challenges for implementation, many scientists remain circumspect, especially with regards to cancer treatments. “At some point, we either need to revisit chronotherapy completely and put in some effort to get more data and make this work,” says Grothey, “or we say, ‘OK, that was just a side note in the history of oncology.’” He recalls fleeting excitement surrounding chronotherapy when he entered the cancer field about 20 years ago. “A lot of us discarded it as something that was too complicated,” he says. “We didn't have the technology. That might be different now.” ■

Lynne Peeples is a science journalist in Seattle, Washington. Additional reporting in French provided by Sabine Louet.

1. Slat, E. A. *et al. J. Biol. Rhythms* **32**, 121–129 (2017).
2. Kobayashi, M., Wood, P. A. & Hrushesky, W. J. M. *Chronobiol. Int.* **19**, 237–251 (2002).
3. Giacchetti, S. *et al. Cancer Res.* **77** (Suppl. 4, abstr. P4-06-06) (2017).
4. Li, J., Chen R., Ji, M., Zou, S. L. & Zhu, L. N. *Cancer Chemother. Pharmacol.* **76**, 651–655 (2015).
5. Selfridge, J. M. *et al. Drugs* **76**, 1507–1521 (2016).
6. Pitter, K. L. *et al. Brain* **139**, 1458–1471 (2016).
7. Leder, K. *Cell* **156**, 603–616 (2014).
8. Lévi, F., Zidani, R. & Misset, J.-L. *Lancet* **350**, 681–686 (1997).
9. Montaigne, D. *et al. Lancet* **391**, 59–69 (2018).
10. Chan, S. *et al. Clin. Oncol. (R. Coll. Radiol.)* **29**, 231–238 (2017).
11. Giacchetti, S. *et al. J. Clin. Oncol.* **24**, 3562–3569 (2006).
12. Fischer, D., Lombardi, D. A., Marucci-Wellman, H. & Roenneberg, T. *PLOS ONE* **12**, e0178782 (2017).
13. Chen, C.-Y. *et al. Proc. Natl Acad. Sci. USA* **113**, 206–211 (2016).
14. Hermida, R. C., Ayala, D. E., Fernández, J. R., Mojón, A. & Smolensky, M. H. *Eur. J. Clin. Invest.* <https://doi.org/10.1111/eci.12909> (2018).

COMMENT

CYBERATTACKS Regulate artificial intelligence to avert an arms race in cyberspace **p.296**

ECONOMICS Redefine value to reward innovators, not corporations **p.300**

AWARDS Why the physics Nobel needs an overhaul **p.301**

NATURAL HISTORY The museum heist that became a trail of feathers **p.303**



JACK TAYLOR/GETTY



Police collect samples in Salisbury, UK, following the poisoning of Sergei and Yulia Skripal with a nerve agent in March.

How to curb production of chemical weapons

Companies that manufacture and distribute the precursors to lethal agents must be open to surveillance and inspections, argues **Leiv K. Sydnès**.

Incidents involving chemical weapons are on the rise. In the past week, reports from Syria allege that scores of people in the city of Douma were killed with a toxic gas, possibly chlorine, a tactic that experts say has been used in Syria at least a dozen times since 2012 (see go.nature.com/2hjzc20). Last month in Salisbury, UK, former Russian intelligence officer Sergei Skripal, his daughter

Yulia and a police officer were exposed to an organophosphate called novichok, one of a family of nerve agents said to be the deadliest known¹. And Kim Jong-nam, the eldest son of former North Korean leader Kim Jong-il, was assassinated in 2017 through exposure to another nerve agent, VX, at Kuala Lumpur international airport in Malaysia.

These recent events risk reversing two

decades of progress in disarmament. The intention of the Chemical Weapons Convention (CWC)², finalized in 1992, was to free the world of this weaponry. The Organization for the Prohibition of Chemical Weapons (OPCW), which has implemented the convention since 1997, aimed to destroy all declared stockpiles of chemical weapons within a decade. That hasn't happened, ►

▶ but it is still within reach. Today, 96% of known stockpiles have gone.

It is crucial that, in uncertain times, nations do not fall back on using chemical weapons. In the past few years, political positions and structures that served people well for decades have been questioned. Insecure countries might become more willing to apply chemical weapons to harm opponents and secure strategic advantages.

Meanwhile, the OPCW has lost some of its bite. Although investigators have been allowed in, Syria's breaches have gone unpunished. And neither the United Kingdom nor Malaysia called in experts from the organization right away to deal with their recent nerve-agent attacks, even though immediate assistance is available under the convention.

A new mindset is needed. It is impossible to ban every chemical that could be used to make a weapon, because almost all of them have other applications. For example, chlorine is a common industrial reagent as well as a suffocating gas. More than 60 million tonnes are produced each year and used for purifying water and manufacturing plastics, solvents and pharmaceuticals. Organophosphates are the basis of insecticides and herbicides as well as precursors of nerve agents. Many deadly compounds are easy for any professional chemist to make, with access to the right materials.

There are two solutions: monitor the production and distribution of certain key chemicals (such as organophosphates) that might be misused; and train chemists to be aware of potential security risks.

To realize both, the OPCW should be strengthened and revised. The organization's mandate should be expanded to monitor closely the production of the precursor chemicals used to make the deadliest weapons, especially nerve agents. Its experts must lead all investigations of incidents involving such agents.

Meanwhile, chemists in industry and academia must sign up to a code of conduct surrounding the production, sale and use of chemicals, especially those listed in the CWC. Each time a chemical weapon is used, the reputation of chemists and the chemicals industry is imperilled.

DUAL USES

Almost any chemical can, in principle, be used as a weapon. Most are inconvenient if the goal is to kill or frighten lots of people quickly. But many can be misappropriated.

Chemical weapons fall broadly into three groups: poisonous commodity chemicals, mustard compounds and nerve agents (see 'Classes of chemical weapon'). Each kills in a different way, by blocking or triggering reactions in the body. The signatures left behind in a person's tissues differ for each weapon and can reveal which compound

was used. Nerve agents are relatively easy to trace, compared to a chlorine attack. They form relatively stable molecular products by linking to biomolecules such as proteins, which can then be sampled in tissue, serum and urine. These products can be converted into other compounds that reveal which agent was used.

The chemical analyses required for such detective work are highly specialized. They demand skilled personnel who know how to prepare the samples and which safety measures and precautions to apply. Detailed measurements must be made with sophisticated instruments, such as gas or liquid chromatographs and mass spectrometers.

Around two dozen laboratories worldwide are capable of doing this forensic work^{3,4}. The OPCW carries out quality-control tests on such labs twice a year and accredits those that pass. Almost all are government facilities that have ties to the military. For example, the Skripal samples were sent to the UK government's Defence Science and Technology Laboratory (DSTL) at Porton Down.

In my view, the OPCW — and no one else — should take charge of all cases involving chemical weapons, in particular

"No one can usually say for sure who launched or dropped the shell carrying the chemical weapon."

suspected nerve agents and mustard compounds. The application of such weapons by any state party to the CWC is an outrageous breach. The implications are so serious and delicate that an impartial, experienced body must resolve the situation. The OPCW is such a body. It should control every step of the process, from the collection and storage of samples to the release of the results. This way, improper interventions will be impossible and the results can be widely trusted.

Regrettably, this procedure was not followed in the Kim and Skripal cases, which remain politically fraught. The Malaysian and UK governments reacted to the incidents on their own, despite the OPCW being able to initiate a rapid response within 24 hours. The UK government officially informed the OPCW secretariat four days after the Skripal incident, asking it to independently verify the novichok structure; the OPCW has now done so and supports the United Kingdom's assessment. I believe that it would have been better to have had international oversight of the samples from the outset.

Another virtue of the OPCW is its sobriety when presenting the results of analyses. It gives only the evidence that proves which agent it has found; it does not speculate as to which party was behind a violation. Such restraint is exercised for good reasons. No one can usually say for sure who launched or dropped the shell carrying the chemical

CLASSES OF CHEMICAL WEAPON

Three fatal mechanisms

Commodity chemicals. Industrial chemicals used as weapons include chlorine (Cl_2), phosgene ($\text{Cl}_2\text{C}=\text{O}$) and hydrogen cyanide (HCN). Chlorine, a pale green gas heavier than air, suffocates people and destroys the lungs. Its use is hard to prove directly. Samples of the gas must be collected within minutes, before it disperses. Eyewitness interviews and medical records are used instead. The first large-scale use of chlorine in war was on 22 April 1915. The German army released 168 tonnes from 6,000 barrels near Ypres, Belgium. (The chlorine programme's leader was Fritz Haber, who is better known for his work on ammonia synthesis.)

Mustard compounds. These organic compounds contain sulfur or nitrogen groups (such as bis(2-chloroethyl) sulfide ($\text{ClCH}_2\text{CH}_2\text{SCH}_2\text{CH}_2\text{Cl}$)) that react with skin and the airways. Mustard gas, a colourless liquid that boils at 217°C , has been used in warfare since 1917 and causes

swelling and blisters. The type of agent used can be determined by the presence of stable biomarkers in biopsy samples. A mustard compound was used during the Iran–Iraq war in Halabja in the Kurdistan autonomous region of Iraq on 16 March 1988: 5,000 people died immediately and many more were injured.

Nerve agents. These phosphorus-based organic compounds affect the transmission of signals across nerve junctions. They inhibit the enzyme acetylcholinesterase, and so increase the amount of one type of nerve signal that reaches muscles, causing paralysis. The agents are colourless liquids that can kill within a minute. People exposed convulse and foam at the mouth, then their respiratory system and heart muscles fail. Molecules indicating which agent was used can be collected from tissues. One example is the release of sarin on the Tokyo subway in 1995 by a cult group, in which 12 people were killed. [L.K.S.](#)





Disabled fermentation vats and equipment at the Muthanna chemical-weapons factory in Iraq.

SCOTT PETERSON/GETTY

(all sides would deny it). And it is impossible to tell where an agent was produced without also having samples from that same location with which to match it. With no comparison samples, it is like trying to identify a person from only a fingerprint without access to a database of prints.

In the UK novichok case, as far as I know, the precursors and the reagents used to prepare this nerve agent are unavailable to cross-check. If that is so, then the origin of that weapon could not be established on the basis of the chemistry alone — other evidence or intelligence would be needed to determine the chemical's source. Gary Aitkenhead, chief executive of the DSTL, clarified this point a month after the incident.

Nor does the fact that Russian chemists first developed novichok compounds indicate definitively where this material came from. Information about novichok nerve agents has been available in the literature for years^{5–7}. In theory, it is not technically difficult to synthesize novichok compounds, although some are harder to make than others. The difficulties lie in obtaining the necessary precursors and reagents, and safely preparing the agent without self-exposure.

PREVENT PROLIFERATION

To reiterate, we need to do two things to limit chemical weapons: control the ingredients and improve ethical standards in the chemical profession.

Banning chemicals is impossible because almost all the relevant chemicals required for making chemical weapons have good uses as well as bad. For example, isopropyl alcohol (IPA) is widely used as a solvent; millions of tonnes are used each year in the production of a range of products, including household

cleaners, pesticides and personal-care products. Yet react IPA with methylphosphonyl difluoride, and you produce the nerve agent sarin.

The focus, therefore, must be on tightening security around particular commercial chemicals. The priority should be surveillance and inspection of the producers and distributors of the relatively few phosphorus compounds that can be used to make nerve agents (including the five groups of precursors and seven specific precursors listed in the CWC). The OPCW must be in charge of this inspectorate, because it would require the backing of state parties to the CWC.

Meanwhile, the chemistry community — in both academia and industry — needs to become more aware of the potential misuse of certain chemicals. For example, in the 1990s, a colleague of mine inspected the main Iraqi facility that produced chemical weapons in the 1980s — the Muthanna State Establishment, northwest of Baghdad. To his disbelief, he discovered 4,000 tonnes of weapons ready to be launched (mainly mustard gas and cyclosarin), as well as 20,000 tonnes of precursor chemicals for making them. Storerooms were filled with barrels of thionyl chloride — an industrial chemical listed under the CWC as having possible weapons uses. Many of these were bought from European companies⁸. Presumably no red flags were raised among company staff when the sizes of the Iraqi orders went from kilograms to tonnes.

In my experience, few chemists know which chemicals to pay attention to from a weapons perspective. Most have never heard of the three lists of precursors and toxic chemicals (schedules 1–3) in the CWC,

even if they are aware of the convention itself. Hardly any universities incorporate weapons-related topics in curricula for chemists and chemical engineers. This has to change. And here the OPCW can assist, through ethical guidance and educational material.

In 2015, after a lengthy process, the OPCW took a leap by publishing The Hague Ethical Guidelines for chemistry professionals (see go.nature.com/2epdgrj). It also set up an Advisory Board on Education and Outreach, which has begun to post educational materials online (see go.nature.com/2jnzmp9). So far, it has had little impact. More needs to be done to spread the word.

EXPAND MANDATE

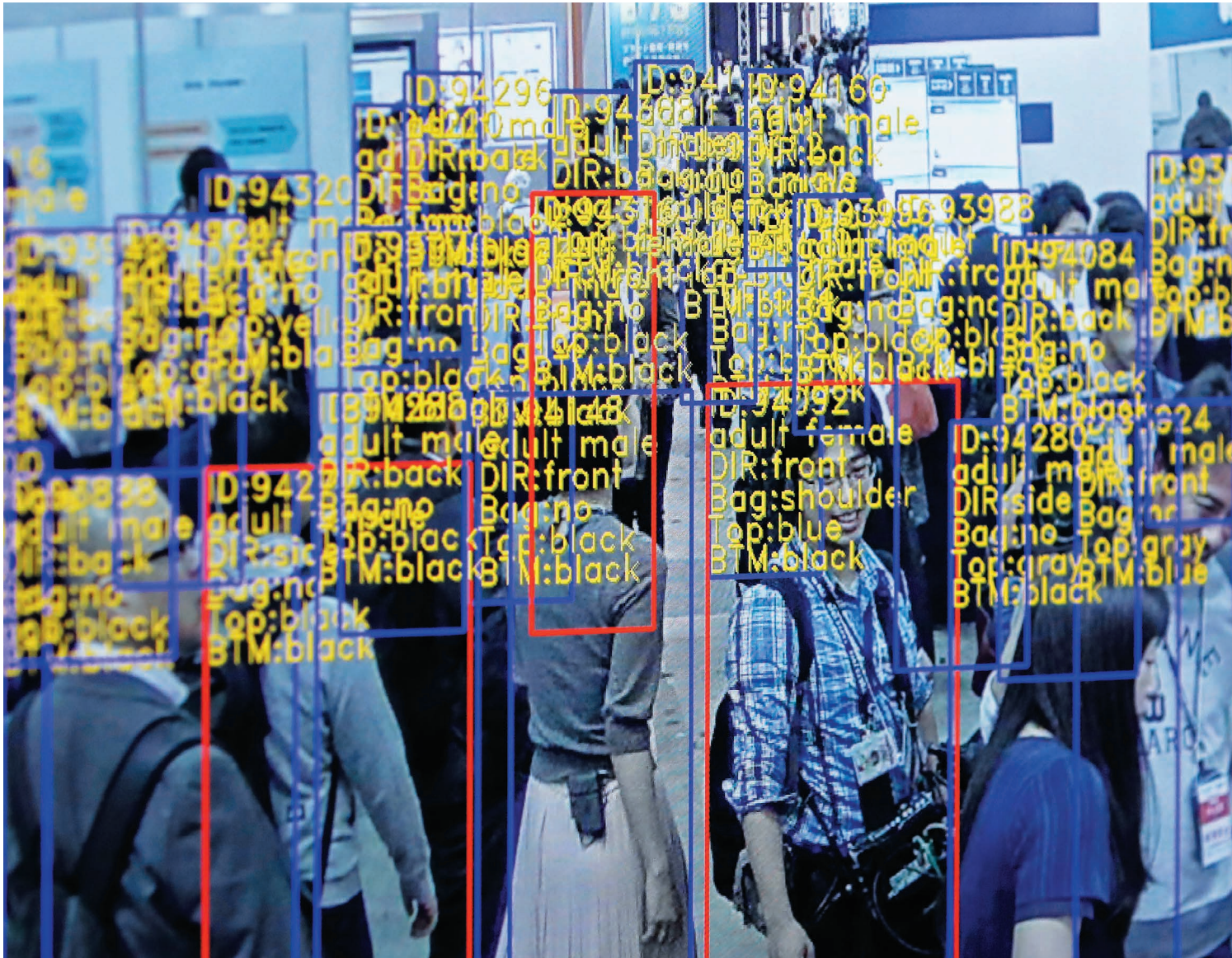
To stop renewed proliferation, the OPCW's powers, roles and influence should be expanded, so that it can act more quickly and forcefully when the CWC is breached or other threats arise. State parties to the convention should enable the OPCW to become more heavily involved in awareness-raising, inspections, outreach and surveillance. And the permanent delegations to the OPCW should contact universities and professional organizations in their countries to highlight these important issues among chemists and within the chemical industry.

Research chemists, especially in universities, should work to raise awareness of the chemical challenges related to the CWC. A first step would be to make the convention mandatory reading for all chemistry students. Second, OPCW educational material should be used in university courses. Third, The Hague Ethical Guidelines should be used to improve the ethical framework of the chemical profession. ■

Leiv K. Sydnès is professor of chemistry at the University of Bergen, Norway. He chaired the international task group that assessed the impact of scientific advances on the Chemical Weapons Convention in 2007 and 2012.

e-mail: leiv.sydnès@uib.no

1. Malygin, V. V., Sokolov, V. B., Richardson, R. J. & Makahaeva, G. F. *J. Toxicol. Environ. Health A* **66**, 611–625 (2003).
2. Organisation for the Prohibition of Chemical Weapons (OPCW). *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction* (OPCW, 2005).
3. OPCW Technical Secretariat. *Status of the Laboratories Designated for the Analysis of Authentic Biomedical Samples* (OPCW, 2017).
4. OPCW Technical Secretariat. *Status of the Laboratories Designated for the Analysis of Authentic Environmental Samples* (OPCW, 2017).
5. Hoenig, S. L. *Compendium of Chemical Warfare Agents* (Springer, 2007).
6. Kruglyak, Y. L., Malekin, S. I. & Martynov, I. V. *Zh. Obs. Khim.* **42**, 811–814 (1972).
7. Raevskii, O. A. et al. *Zh. Obs. Khim.* **57**, 2073–2078 (1987).
8. Pearson, G. S. *The Search for Iraq's Weapons of Mass Destruction: Inspection, Verification and Non-Proliferation* (Palgrave/Macmillan, 2005).



Object detection and tracking technology on show at the 2018 Artificial Intelligence Exhibition & Conference in Tokyo.

Regulate artificial intelligence to avert cyber arms race

Define an international doctrine for cyberspace skirmishes before they escalate into conventional warfare, urge **Mariarosaria Taddeo** and **Luciano Floridi**.

Cyberattacks are becoming more frequent, sophisticated and destructive. Each day in 2017, the United States suffered, on average, more than 4,000 ransomware attacks, which encrypt computer files until the owner pays to release them¹. In 2015, the daily average was just 1,000. In May last year, when the WannaCry virus crippled hundreds of IT systems across the UK National Health Service, more

than 19,000 appointments were cancelled. A month later, the NotPetya ransomware cost pharmaceutical giant Merck, shipping firm Maersk and logistics company FedEx around US\$300 million each. Global damages from cyberattacks totalled \$5 billion in 2017 and may reach \$6 trillion a year by 2021 (see go.nature.com/2gncsyg).

Countries are partly behind this rise. They use cyberattacks both offensively and

defensively. For example, North Korea has been linked to WannaCry, and Russia to NotPetya.

As the threats escalate, so do defence tactics. Since 2012, the United States has used 'active' cyberdefence strategies, in which computer experts neutralize or distract viruses with decoy targets, or break into a hacker's computer to delete data or destroy the system. In 2016, the United Kingdom announced a



KIYOSHI OTA/BLOOMBERG/GETTY

5-year, £1.9-billion (US\$2.7-billion) plan to combat cyber threats. NATO also began drafting principles for active cyberdefence, to be agreed by 2019. The United States and the United Kingdom are leading this initiative. Denmark, Germany, the Netherlands, Norway and Spain are also involved (see go.nature.com/2hebxnt).

Artificial intelligence (AI) is poised to revolutionize this activity. Attacks and responses will become faster, more precise and more disruptive. Threats will be dealt with in hours, not days or weeks. AI is already being used to verify code and identify bugs and vulnerabilities. For example, in April 2017, the software firm DarkTrace in Cambridge, UK, launched Antigena, which uses machine learning to spot abnormal behaviour on an IT network, shut down communications to that part of the system and issue an alert. The value of AI in cybersecurity was \$1 billion in 2016 and is predicted to reach \$18 billion by 2023 (ref. 2).

By the end of this decade, many countries plan to deploy AI for national cyberdefence;

for example, the United States has been evaluating the use of autonomous defence systems and is expected to issue a report on its strategy next month³. AI makes deterrence possible because attacks can be punished⁴. Algorithms can identify the source and neutralize it without having to identify the actor behind it. Currently, countries hesitate to push back because they are unsure who is responsible, given that campaigns may be waged through third-party computers and often use common software.

The risk is a cyber arms race⁵. As states use increasingly aggressive AI-driven strategies, opponents will respond ever more fiercely. Such a vicious cycle might lead ultimately to a physical attack.

Cyberspace is a domain of warfare, and AI is a new defence capability. Regulations are thus necessary for state use of AI, as they are for other military domains — air, sea, land and space⁶. Criteria are needed to determine proportional responses, as well as to set clear thresholds or ‘red lines’ for distinguishing legal and illegal cyberattacks, and to apply appropriate sanctions for illegal acts⁷. In each case, unilateral approaches will be ineffective. Rather, an international doctrine must be defined for state action in cyberspace. Alarming, international efforts to regulate cyber conflicts have stalled.

We call on regional forums, such as NATO and the European Union, to revive efforts and prepare the ground for an initiative led by the United Nations. In the meantime, computer experts must be transparent about problems, limitations and shortcomings of using AI for defence. Researchers must also work with policymakers and end users to design testing and oversight mechanisms for this technology.

NO RULES

Right now, the UN process is in deadlock. In 2004, the UN set up the Group of Governmental Experts on Information Security to agree on voluntary rules for how states should behave in cyberspace. Its fifth meeting, in 2017, ended in a stand-off. The group could not reach consensus on whether international humanitarian law and existing laws on self-defence and state responsibility should apply in cyberspace. The United States argued that cyberdefence regulations should build on these laws. Other nations, including Cuba, Russia and China, disagreed. They argued that this would ‘militarize’ cyberspace and send the wrong message about peaceful conflict resolution. The group failed to deliver its report. It is unclear whether it will meet again, or what will happen next.

International dialogue and action must

resume. NATO could pave the way through its forthcoming guidelines, although it is currently unclear what their scope will be.

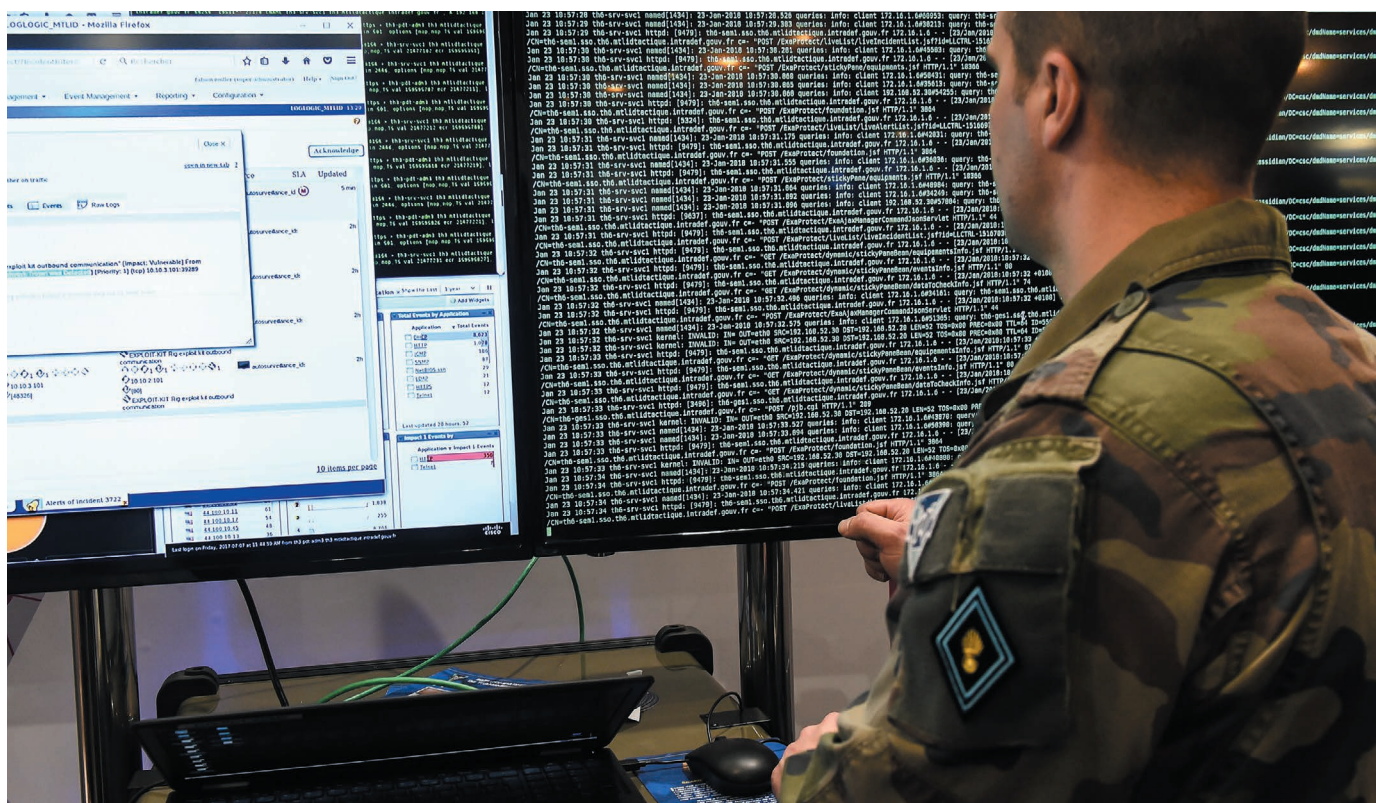
Meanwhile, research on AI for cyberdefence is progressing quickly. The United States is in the lead, technologically. It aims to incorporate AI into its cyberdefence systems by 2019 (ref. 3). The US Department of Defense (DOD) has earmarked \$150 million for research. The US Defense Advanced Research Projects Agency (DARPA) is developing the techniques and strategies. Steps have already been taken. In DARPA’s 2016 Cyber Grand Challenge competition, seven AI systems, developed by teams from the United States and Switzerland, fought against each other. The systems identified and targeted their opponents’ weaknesses while finding and patching their own.

The DOD will issue the first US report on AI strategies for national defence in May. There is, as far as we know, no indication of what its approach will be. Previous documents, such as *The DOD Cyber Strategy* from 2015 or the 2016 *National Cyber Incident Response Plan*, did not cover autonomous systems, machine learning or AI. The 2012 DOD directive on ‘Autonomy in Weapon Systems’ focused on internal procedures for deploying AI but was silent on when the United States would do so in the international arena.

AI is a priority for China, which aims to become a world leader in machine-learning technologies. In July 2017, the Chinese government issued its Next Generation AI Development Plan. Military implementation of AI, on the battlefield as well as in cyberspace, is a crucial part of the strategy. But it is unclear to what degree China plans to deploy AI actively in cyberdefence.

Russia has not released any public documents about its strategies for AI in defence. However, in a video message released in 2017, President Vladimir Putin referred to AI and stated: “Whoever becomes the leader in this sphere will become the ruler of the world.” Experts agree that Russia is focusing on developing AI-enhanced tools for its conventional forces. However, since 2014, the Russian National Defense Control Center has been using machine-learning algorithms to detect online threats. Allegedly, Russia has pioneered the use of AI to spread disinformation and intervene in the public debates of other nations, including the 2016 US presidential election and the United Kingdom’s EU membership referendum. Although these operations are not part of national defence strategies, they indicate Russia’s advanced AI capabilities.

North Korea has a history of cyberspace aggression. It was implicated, for example, in the WannaCry attack in 2016 and in another major breach, against Sony Pictures, in 2014. The country lacks technical expertise in AI but is likely to want to catch up with its adversaries.



A military cyberdefence specialist at a conference in Lille, France. Government spending on cyber strategies has soared over the past decade.

The EU is stepping up, too. In 2017, it reassessed cybersecurity and defence policies and launched the European Centre of Excellence for Countering Hybrid Threats, based in Helsinki. The EU has the most comprehensive regulatory framework for state conduct in cyberspace so far. Yet these directives do not go far enough. The EU treats cyberdefence as a case of cybersecurity, to be improved passively by making member states' information systems more resilient. It disregards active uses of cyberdefence and does not include AI.

This is a missed opportunity. The EU could have begun defining red lines and proportionate responses in its latest rethink. For example, the 2016 EU directive on 'Security of Network and Information Systems' provides criteria for identifying crucial national infrastructures, such as health systems or key energy and water supplies that should be protected. The same criteria could be used to define illegitimate targets of state-sponsored cyberattacks.

Regional forums, such as NATO and the EU, must take the following three steps to avoid serious imminent attacks on state infrastructures, and to maintain international stability.

THREE STEPS

Define legal boundaries. The international community needs to agree urgently on red lines that distinguish between legitimate and illegitimate targets. Also needed are definitions of proportionate responses

for cyberdefence strategies. International consensus at the UN level will ultimately be required. Until then, guidelines from regional multilateral bodies, such as NATO and the EU, must cover these issues and lead by example.

Test strategies with allies. 'Sparring' exercises should be organized between friendly countries to test AI-based defence tactics. These tests should be mandatory before any system is deployed. They could be in the form of DARPA's Grand Challenge

"The EU could have begun defining red lines and proportionate responses in its latest rethink."

and the simulation exercises routinely run by NATO and the EU. Because AI learns by experience, these matches will improve the strategies of the alliance, while finding and healing weaknesses. Fatal vulnerabilities of key systems and crucial infrastructures should be shared with allies; policy frameworks should demand disclosure. Agreements and regulations with similar sharing and disclosure requirements include the EU Electronic Identification, Authentication and Trust Services Regulation and NATO's Industry Partnership Agreement.

Monitor and enforce rules. The international community needs to agree how to audit and oversee AI-based state cyberdefence operations. 'Alert and remedy'

mechanisms are needed to address mistakes and unintended consequences. A third-party authority with teeth, such as the UN Security Council, should rule on whether red lines, proportionality, responsible deployment or disclosure norms have been breached. Economic or political sanctions should be imposed on states that violate rules. NATO and the EU should enforce the norms within their remits.

The solution is difficult, but it is clear. There is no time to waste. ■

Mariarosaria Taddeo is a research fellow and deputy director of the Digital Ethics Lab at the Oxford Internet Institute, University of Oxford, UK; and a Turing fellow of the Alan Turing Institute, London, UK. **Luciano Floridi** is professor of philosophy and ethics of information at the University of Oxford, UK; director of the Digital Ethics Lab at the Oxford Internet Institute; and chair of the Data Ethics Group at the Alan Turing Institute.
e-mail: mariarosaria.taddeo@oii.ox.ac.uk

1. US FBI. *How to Protect Your Networks from Ransomware* (FBI, 2017).
2. P&S Market Research. *Artificial Intelligence (AI) in Cyber Security Market* (P&S Market Research, 2017).
3. US Defense Science Board. *Summer Study on Autonomy* (US Department of Defense, 2016).
4. Taddeo, M. *Phil. Technol.* <https://doi.org/10.1007/s13347-017-0290-2> (2017).
5. Yang, G.-Z. *et al. Sci. Robot.* **3**, eaar7650 (2018).
6. Floridi, L. *Phil. Trans. R. Soc. A* **374**, 20160112 (2016).
7. Taddeo, M. *Minds Mach.* **27**, 387–392 (2017).

SPRING BOOKS



ECONOMICS

The reinvention of value

Robert Costanza applauds Mariana Mazzucato's call for more-productive, equal and sustainable economies.

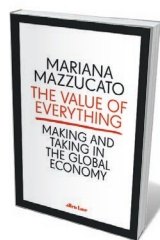
What we value and how we value it is one of the most contested, misunderstood and important ideas in economics. Economist Mariana Mazzucato's comprehensive *The Value of Everything* explores how ideas about what value is, where it comes from and how it should be distributed have changed in the past 400 years, and why value matters now more than ever. Mazzucato emphasizes the need to reopen debate to make economies more productive, equitable and sustainable. The 2008 financial crisis was just a taste of looming problems — climate disruption, massive biodiversity and ecosystem-services decline, even the possible collapse of Western civilization — unless we learn to value what really matters.

Early economists focused on the production of value from land (François Quesnay and the 'physiocrats'), labour (Adam Smith to Karl Marx) and capital. In this view, value determines price (Four decades ago, I described this in terms of embodied energy: see R. Costanza *Science* **210**, 1219–1224; 1980). By contrast, the current mainstream

'marginalist' concept bases value on market exchanges: price, as revealed by the interaction of supply and demand in markets, determines value, and the only things that have value are those that fetch a price.

This has major implications for ideas about the distinction between value creation and value extraction, the nature of unearned income ('rent') and how value should be distributed. As Mazzucato notes, it stokes inequality because the market, simply by generating income, is seen to justify its level and distribution: "All income, according to this logic, is earned income: gone is any analysis of activities in terms of whether they are productive or unproductive."

Mazzucato lays out disturbing implications



The Value of Everything: Making and Taking in the Global Economy
MARIANA MAZZUCATO
Allen Lane (2018)

of the marginalist approach. These include (mis)measuring national income and real wealth, confusing financial speculation with the production of value, perverting the patent system (which stifles, rather than rewards, innovation) and undervaluing government and public goods, including public infrastructure, ecosystems and social networks. Her engaging and insightful exploration reveals how embedded the marginalist approach has become, and how it distorts economies' ability to foster innovation, equity and real progress.

The international System of National Accounts and gross domestic product (GDP) both value economic activity on the basis of market transactions — only goods and services sold in markets are counted. Much of that activity is beneficial, but some is best seen as a cost to be avoided. GDP conflates the two. For instance, growth of crime demands more police and security devices; these add to GDP, but more crime is not desirable. Increases in air and water pollution, serious illness and divorce are all counted as positive in GDP, whereas the distribution of income is ignored,

ILLUSTRATIONS BY STEPHAN SCHMITZ

as are the value of household and volunteer work, ecosystem services and community support. As economist and statistician Simon Kuznets, GDP's main architect, warned, a country's welfare cannot be inferred from GDP: "Goals for more growth should specify more growth of what and for what."

Mazzucato argues persuasively that GDP is a "hodge-podge" that "invites lobbying rather than reasoning about value". She notes that it "justifies excessive inequalities of income and wealth and turns value extraction into value creation". One alternative measure is the Genuine Progress Indicator (GPI), which attempts to separate environmental and social costs from benefits, to value household and volunteer work, and to adjust for inequality. For many countries, including the United States, China and the United Kingdom, there have been no net gains in GPI for several decades (I. Kubiszewski *et al. Ecol. Econ.* **93**, 57–68; 2013). You get what you measure, and misusing GDP as a policy goal is distorting decisions about real progress (R. Costanza *et al. Nature* **505**, 283–285; 2014).

Mazzucato deconstructs several other key trends. These include how the financial sector's "casino capitalism" mislabels market speculation as the creation of value rather than the mere extraction of value created elsewhere, and how the real value added by government and public goods and services have been ignored — to the detriment of us all. Ultimately, she notes, we need a more synthetic and integrative view: one that recognizes both how value is created and extracted in the current system, and how this needs to change. She concludes that value depends on vision: "If we cannot dream of a better future and try to make it happen, there is no real reason why we should care about value." The ability to value a healthy, sustainable planet, fairness, community and quality of life must be returned to the heart of economics.

Economics has been defined as the use of scarce resources to achieve desirable ends. In the Anthropocene epoch of human influence on the planet, we need to redefine those ends, and reevaluate which resources are truly scarce. Value should be viewed as contribution to the sustainable well-being of Earth and all its inhabitants. The United Nations Sustainable Development Goals are a huge step towards a broad global consensus on a desirable economy and society. As US baseball player Yogi Berra quipped: "If you don't know where you're going, you'll end up somewhere else." Mazzucato's trenchant analysis is a compelling call to reinvent value as a key concept to help us achieve the world we all want. ■

Robert Costanza is a professor of ecological economics and Vice-Chancellor's Chair in Public Policy at the Crawford School of Public Policy of the Australian National University in Canberra.
e-mail: robert.costanza@anu.edu.au

COSMOLOGY

The trouble with the Nobel prize

Ron Cowen weighs up Brian Keating's call to reform the most coveted award in physics.

If cosmologist Brian Keating had his way, the scientific teams that made two of the most astounding discoveries in physics — the Higgs boson and gravitational waves — would never have won Nobel prizes.

It's not that Keating thinks the researchers undeserving. But the current rules and structure of the awards, he contends in *Losing the Nobel Prize*, foster ferocious and sometimes destructive competition for scarce research resources. He avers that the prizes are also biased against the work of female and younger scientists, and that they violate some of the very principles that Alfred Nobel, their founder, specified in his will more than a century ago.

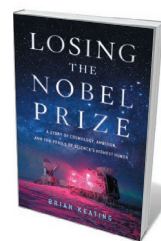
Keating studies the infant Universe through subtle patterns in the cosmic microwave background (CMB) left over from the Big Bang. He is a deft writer, interweaving the science with personal musings on topics from his relationship with a father who abandoned him as a child to the passions that impel him to explore the unknown. Looming over all are his concerns about the Nobels.

These arose after his very public roller-coaster ride as part of a research team whose work briefly seemed a shoo-in for the physics prize. The team — a collaboration between institutions including the Harvard-Smithsonian Center for Astrophysics (CfA) in Cambridge, Massachusetts, and the University of California, San Diego (UCSD) — had built two radio telescopes at the South Pole to hunt for a signature in the CMB that could reveal how the early Universe had evolved. Keating conceived the first, BICEP1. The team then developed the more sensitive BICEP2, which observed the CMB from 2010 to 2012.

Rumours of a scientific coup began flying in March 2014, even before the CfA

alerted the media of an imminent "major discovery". The press briefing on 17 March did not disappoint (I was there, covering the event for *Nature's* news section). The team's four principal investigators, who included astronomer John Kovac, reported that they had detected a subtle twist in CMB polarization. They asserted that its source was almost

certainly primordial gravitational waves, which would have been generated by inflation — a brief, faster-than-light ballooning of the infant Universe. That theoretical growth spurt had been a cornerstone of



Losing the Nobel Prize: A Story of Cosmology, Ambition, and the Perils of Science's Highest Honor
BRIAN KEATING
W. W. Norton (2018)

cosmology for some 35 years, but definitive proof had not been found.

BICEP2's discovery reverberated across the media. At the briefing, accolades poured in. Keating, one of several team members not there, recounts his mixture of frustration and elation: although Kovac mentioned his work, it was not cited in the press release. Keating well

knew that if a Nobel had been in the offing, he and most of the team would have been excluded, given the focus on principal investigators, and the rule that any prize can be shared by a maximum of three people.

The glory was, in any case, not to be. For months, Keating watched from the sidelines as the discovery literally turned to dust. All along, the BICEP2 team had worried that hydrocarbon soot and other cosmic particles could confound the results. (When light, including the CMB, reflects off non-spherical particles of galactic dust whose ▶

THE CURRENT
RULES FOSTER
FEROCIOUS
AND SOMETIMES
DESTRUCTIVE
COMPETITION
FOR SCARCE RESEARCH
RESOURCES.



► axes are aligned, the light is imprinted with the same curlicue polarization pattern expected from gravitational waves in the early Universe.) Yet the team decided to go ahead with the announcement, buoyed by data from a slide used for a 2013 talk by a scientist affiliated with BICEP2's chief competitor, the European Space Agency's Planck satellite.

The slide showed an unpublished dust map of unknown accuracy. Extrapolating from it, the BICEP2 team concluded that in the region of sky observed by its telescope, galactic dust would have little effect on the results. Keating writes that he objected to relying on such evidence for a high-stakes discovery, but was ultimately swayed. New data from the Planck satellite later revealed that dust had led the BICEP2 team to misread the results. Its vision, Keating feels, had been clouded not only by dust, but by 'Nobel lust' and the fear of being scooped.

Journalists embraced the BICEP2 announcement at first. It was an exhilarating story to report, and I have since debated whether its potential might have clouded my own vision. The dozen or so independent

experts I contacted, who had read advance copies of a paper that the BICEP2 team would later post online, all commented positively on the work. But it's possible that for a few, confirmation bias played a part, because they were proponents of the inflation theory.

Keating suggests several remedies for Nobel fever. He argues that the physics prizes should be awarded only for serendipitous findings; an example is the evidence, discovered in 1998 by two teams of cosmologists, that the Universe was revving up its expansion instead of slowing down. If a team finds something it had set out to look for, it should not gain the Nobel, is his provocative view. Keating also asserts that Nobel prizes should be awarded to an entire team. He would eliminate the stipulation, added in 1974, that the prizes cannot be awarded posthumously. And he would allow more than one prize for the same research if a person was originally overlooked or ignored (which has, historically, often occurred to women, such as co-discoverer of radio pulsars Jocelyn Bell Burnell).

These changes, he argues, might motivate physicists to think outside the box in conducting research, and might discourage

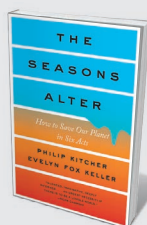
in-fighting. However, I doubt that reconfiguring the Nobels would accomplish what Keating hopes. As he himself notes, both the US and European processes for allocating funding and tenure encourage cut-throat competition. Modifying those ingrained systems would have much greater impact.

Keating notes that his own work has begun to embrace the spirit of cooperation. In 2016, the Simons Foundation, a private philanthropic foundation in New York City that supports research in maths and the basic sciences, gave the green light for him to spearhead a collaboration between his CMB team, based at UCSD, and one based at Princeton University in New Jersey. Together, they hope to dig from the dust a true signal of primordial gravitational waves in the CMB. Even if that pans out, the work would not be eligible for a Nobel under Keating's reforms; it would be science for science's sake. And maybe that's the point. ■

Ron Cowen is a freelance science writer, focusing on physics, astronomy and the history of technology.
e-mail: roncowen@msn.com

NEW IN PAPERBACK

Highlights of this season's releases.



The Seasons Alter: How to Save Our Planet in Six Acts

Philip Kitcher and Evelyn Fox Keller LIVERIGHT (2018)

"Clearly, we need to talk." Philosophers of science Philip Kitcher and Evelyn Fox Keller call for constructive discourse on climate change in their unusual exploration of this urgent, highly politicized issue. While coherently explaining the science, they use Socratic dialogue to explore differing viewpoints. As they warn, considerate, productive conversation is essential if we're not to go down in history as "the people who argued while the world burned".

NATURAL HISTORY

Of plumes and plunder

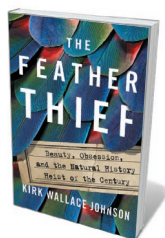
Stuart Pimm on the tragic tale of a hobbyist, a heist and a natural-history collection.

I'll never forget my first sight of a bird of paradise. On a hike in the central highlands of New Guinea one morning, I found an exquisite blue *Paradisaea rudolphi*, hanging upside down, puffing up its feathers, swaying and calling loudly. The birds' very names allude to the sixteenth-century assumption that their exotic feathers — and the lack of feet on skins that made it to Europe — showed that they were celestial beings. Founder of taxonomy Carl Linnaeus knew two species. He followed the myth, naming the greater bird-of-paradise *Paradisaea apoda* (*apoda* meaning 'without feet'). Naturalist Alfred Russel Wallace delivered the first scientific accounts of their behaviours in his 1869 book *The Malay Archipelago*. (Fellow Victorian scientist and ornithological artist John Gould named one species *Semioptera wallacii*.) And who can fail to be enthralled by David Attenborough's BBC films of their otherworldly displays?

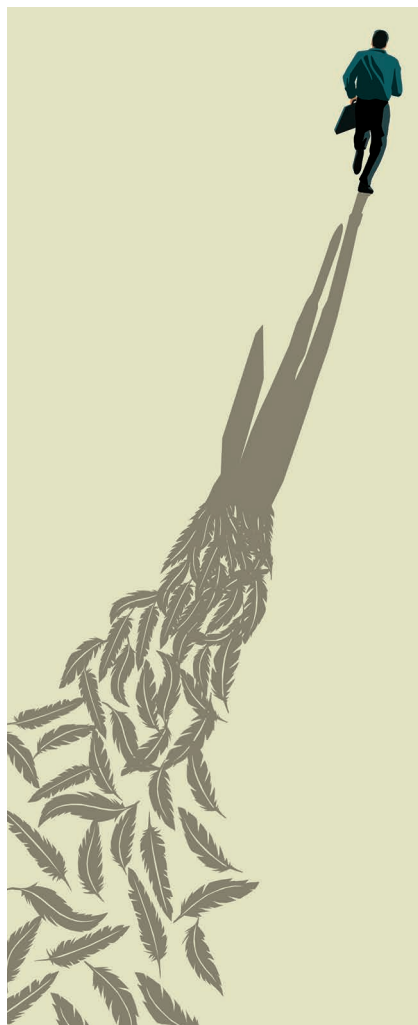
Given all that, it's appalling that museum specimens of these birds — including a number collected by Wallace — were stolen and plucked, and their unique associated data discarded. The culprit was Edwin Rist, practitioner of an arcane art: recreating Victorian ornamental salmon-fishing flies using rare feathers. That heist lies at the core of *The Feather Thief*, in which investigative writer Kirk Wallace Johnson recounts his quest to retrieve what remained of those specimens. Johnson's book also probes how the human yen for the exotic can in some cases harm species and what we know about them.

In June 2009, Rist, a 21-year-old US flautist then studying at the Royal Academy of Music in London, broke into the Natural History Museum's collection at Tring. He came away with 299 stuffed skins of brightly coloured birds, including birds of paradise, showy species of cotingids and quetzals. Rist sold some of the stolen feathers and skins, and used others for his own creations.

Johnson's interest in the story arose from



The Feather Thief: Beauty, Obsession, and the Natural History Heist of the Century
KIRK WALLACE JOHNSON
Viking (2018)

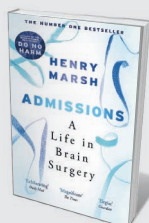


his own devotion to fly fishing. He took up the sport as respite from emotionally draining work helping Iraqis who had worked with US development agency USAID to relocate to the United States. Johnson writes of his fishing reveries: "Out in the river ... five hours would pass in what felt like thirty minutes." One day, his fishing guide told him about Rist and the community devoted to Victorian fly-tying.

When I got to this part of the book, I needed help. My friend David Blinken is a professional fly-fishing guide on Long Island, New York (strictly catch and release). His understanding of fish natural history and behaviour is impressive, exceeded only by his ability to show even me how to catch fish. "Isn't the point of tying a fly to imitate whatever insect is on the water that day? Doesn't entomology matter?" I asked. Blinken replied: "Atlantic salmon are thinking only of reproduction and strike at gaudy objects reflexively. The flies aren't meant to resemble any insect."

I surmised that a salmon fly is so time-consuming to make that it might seem too precious to lose. Blinken said that I had guessed correctly. "Many will never get wet. Most are cherished works of art, enjoyed by a passionate group of collectors at the fringe of the fly-tying community. Attention to historical methods, style and detail are paramount."

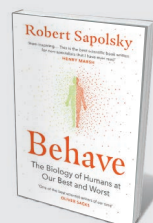
As Johnson reveals, these enthusiasts revived the craft and formed an online community in the late twentieth century. Rist was involved from his early teens, recreating classic flies as a student and then a teacher. One fly might include the 'period-specific' feathers of the golden pheasant (*Chrysolophus pictus*) from China, the red-ruffed fruitcrow (*Pyroderus scutatus*), macaws and the plum-throated cotinga (*Cotinga maynana*) from South America, along with feathers of domesticated birds, such as chickens. But many of the wild birds are rare or endangered, and supply has dwindled to sources such as Victorian feather hats or moulted plumage from zoos. Traded feathers are hugely expensive. ►



Admissions: A Life in Brain Surgery

Henry Marsh WEIDENFELD & NICOLSON (2018)

In this unflinchingly honest memoir, retired neurosurgeon Henry Marsh seamlessly intertwines his life experiences and surgical career. He reflects on both what he has learned by probing the brain, and our limited knowledge of mind, from emotions to consciousness.



Behave: The Biology of Humans at Our Best and Worst

Robert Sapolsky VINTAGE (2018)

Neurobiologist Robert Sapolsky tackles the question of why we behave in the ways we do — whether commendably or despicably. He explores the biology of violence, and examines what it can teach us about altruism.

► (My quick check online showed that pairs from three species of cotinga sell for US\$25 to \$45. I easily found feathers on sale from birds on the International Union for Conservation of Nature's Red List of Threatened Species.)

US thief Willie Sutton allegedly said that he robbed banks because “that’s where the money is”. Museums are where the feathers are. In July 2009, senior curator Mark Adams found the drawers with missing specimens; 16 months later, Rist was caught. He pleaded guilty to burglary and money-laundering. The court fined him £125,150 (US\$200,000 at the time), of which he had about 10%. He also got a mere 12-month suspended sentence, owing to a diagnosis of Asperger’s syndrome (an autism spectrum disorder).

Of the 299 skins stolen, police retrieved only 102 with the labels intact. More had been stripped of the essential data that such labels provide, and 106 were missing. Johnson’s exhaustive sleuthing tracked down some feathers in 2016, but nothing more.

Museum specimens are a unique, contextualized archive, as Robert Prys-Jones, a scientific associate at the Natural History Museum, makes clear in the book. They hold information about where and when species lived, who collected them and perhaps why; and they can be studied for visual and genetic clues. But after interviews with individuals in the fly-tying community, Johnson feels that only some are horrified by the theft. His investigations revealed that the bulk of the birds “dissolved into the bloodstream of the feather underground”, some realms of which seemed to trade in endangered species and flout the Convention on International Trade in Endangered Species of Wild Fauna and Flora. As Blinken told me, the art can become “a pursuit of perfection so intoxicating that its practitioners lose all sense of ethics”.

The Feather Thief is a riveting read. It also stands, I believe, as a reminder of how an obsession with the ornaments of nature — be they feathers, bird eggs or ivory — can wreak havoc on our scientific heritage. ■

Stuart Pimm is professor of conservation at the Nicholas School of the Environment at Duke University in Durham, North Carolina, and directs the non-profit group SavingSpecies, www.savingspecies.org. He tweets at @StuartPimm. e-mail: stuartpimm@me.com



PHYSICS

Stop all the clocks

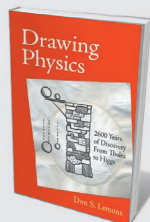
Andrew Jaffe probes Carlo Rovelli’s study arguing that physics deconstructs our sense of time.

According to theoretical physicist Carlo Rovelli, time is an illusion: our naive perception of its flow doesn’t correspond to physical reality. Indeed, as Rovelli argues in *The Order of Time*, much more is illusory, including Isaac Newton’s picture of a universally ticking clock. Even Albert Einstein’s relativistic space-time — an elastic manifold that contorts so that local times differ depending on one’s relative speed or proximity to a mass — is just an effective simplification.

So what does Rovelli think is really going on? He posits that reality is just a complex

network of events onto which we project sequences of past, present and future. The whole Universe obeys the laws of quantum mechanics and thermodynamics, out of which time emerges.

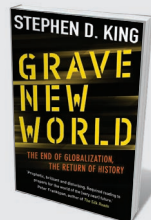
Rovelli is one of the creators and champions of loop quantum gravity theory, one of several ongoing attempts to marry quantum mechanics with general relativity. In contrast to the better-known string theory, loop quantum gravity does not attempt to be a ‘theory of everything’ out of which we can generate all of particle physics and gravitation. Nevertheless, its agenda



Drawing Physics

Don S. Lemons MIT PRESS (2018)

For millennia, drawings have elucidated chewy concepts in physics, providing a “pre-mathematical picture of reality”. Don Lemons delves into the archive for powerful sketches representing ideas and results from Isaac Newton’s colour theory to the Higgs boson.



Grave New World: The End of Globalization, the Return of History

Stephen D. King YALE UNIV. PRESS (2018)

Economist Stephen D. King’s analysis of globalization is searing and timely, offering historical lessons on how political narratives that abandon the global agenda, such as Brexit, threaten our economic order.



of joining up these two fundamentally differing laws is incredibly ambitious.

Alongside and inspired by his work in quantum gravity, Rovelli puts forward the idea of ‘physics without time’. This stems from the fact that some equations of quantum gravity (such as the Wheeler–DeWitt equation, which assigns quantum states to the Universe) can be written without any reference to time at all.

As Rovelli explains, the apparent existence of time — in our perceptions and in physical descriptions, written in the mathematical languages of Newton, Einstein and Erwin Schrödinger — comes not from knowledge, but from ignorance. ‘Forward in time’ is the direction in which entropy increases, and in which we gain information.

The book is split



The Order of Time
CARLO ROVELLI
Allen Lane (2018)

into three parts. In the first, “The Crumbling of Time”, Rovelli attempts to show how established physics theories deconstruct our common-sense ideas. Einstein showed us that time is just a fourth dimension and that there is nothing special about ‘now’; even ‘past’ and ‘future’ are not always well defined. The malleability of space and time mean that two events occurring far apart might even happen in one order when viewed by one observer, and in the opposite order when viewed by another.

Rovelli gives good descriptions of the classical physics of Newton and Ludwig Boltzmann, and of modern physics through the lenses of Einstein and quantum mechanics. There are parallels with thermodynamics and Bayesian probability theory, which both rely on the concept of entropy, and might therefore be used to argue that the flow of time is a subjective feature of the Universe, not an objective part of the physical description.

But I quibble with the details of some of Rovelli’s pronouncements. For example, it is far from certain that space-time is quantized, in the sense of space and time being packaged in minimal lengths or periods (the Planck length or time). Rather, our understanding peters out at those very small intervals for which we need both quantum mechanics and relativity to explain things.

In part two, “The World without Time”, Rovelli puts forward the idea that events (just a word for a given time and location at which something might happen), rather than particles or fields, are the basic constituents of the world. The task of physics is to describe the relationships between those events: as Rovelli notes, “A storm is not a thing, it’s a collection of occurrences.” At our level, each of those events looks like the interaction of particles at a particular position and time; but time and space themselves really only manifest out of their interactions and the web of

causality between them.

In the final section, “The Sources of Time”, Rovelli reconstructs how our illusions have arisen, from aspects of thermodynamics and quantum mechanics. He argues that our perception of time’s flow depends entirely on our inability to see the world in all its detail. Quantum uncertainty means we cannot know the positions and speeds of all the particles in the Universe. If we could, there would be no entropy, and no unravelling of time. Rovelli originated this ‘thermal time hypothesis’ with French

mathematician Alain Connes.

The Order of Time is a compact and elegant book. Each chapter starts with an apt ode from classical Latin poet Horace — I particularly liked “Don’t attempt abstruse calculations”. And the writing, translated from Italian by Erica Segre and Simon Carnell, is more stylish than that in most

physics books. Rovelli ably brings in the thoughts of philosophers Martin Heidegger and Edmund Husserl, sociologist Émile Durkheim and psychologist William James, along with physicist-favourite philosophers such as Hilary Putnam and Willard Van Orman Quine. Occasionally, the writing strays into floweriness. For instance, Rovelli describes his final section as “a fiery magma of ideas, sometimes illuminating, sometimes confusing”.

Ultimately, I’m not sure I buy Rovelli’s ideas, about either loop quantum gravity or the thermal time hypothesis. And this book alone would not give a lay reader enough information to render judgement. *The Order of Time* does, however, raise and explore big issues that are very much alive in modern physics, and are closely related to the way in which we limited beings observe and participate in the world. ■

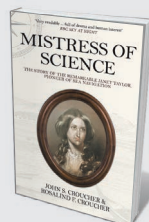
Andrew Jaffe is a cosmologist and head of astrophysics at Imperial College London.
e-mail: a.jaffe@imperial.ac.uk



Life’s Vital Link: The Astonishing Role of the Placenta

Y. W. Loke OXFORD UNIV. PRESS (2018)

This exploration of the placenta’s evolution devotedly details the ‘forgotten’ organ’s vital role in the womb, and other complex functions. Immunologist Y. W. Loke also ponders how such findings could provide insight into his field.



Mistress of Science

John S. Croucher and Rosalind F. Croucher AMBERLEY (2018)

Nineteenth-century British mathematician Janet Taylor has been overlooked by history, yet she invented navigational tools such as the mariner’s calculator, founded an academy and authored textbooks. A fitting tribute to a gifted trailblazer.



PSYCHOLOGY

War and peace and summer camp

Alex Haslam appraises an account of key psychology experiments on conflict and cooperation.

A few years after the Second World War, Muzafer Sherif conducted possibly the most complex field studies ever attempted in social psychology. Sited in summer camps around the United States, they focused on conflict and cooperation within and between two groups of about a dozen 11- and 12-year-old boys. The children were never informed that they were taking part in research. In each study, Sherif and his fellow researchers spent up to three weeks disguised as counsellors and caretakers, manipulating features of the

camp set-up — in particular, the structure of team competitions and challenges — to examine their impact on group relations.

In *The Lost Boys*, Gina Perry puts these extraordinary experiments under the microscope. As in her 2013 book *Behind the Shock Machine*, which probed psychologist Stanley Milgram's 1960s research on obedience, she is unsatisfied with the half-truths lazily handed down in textbooks. Her aim is to make a distinctive contribution to the current debate about replication and reproducibility in social psychology. She goes

in search of the stories behind the research, in particular reassessing Sherif's legacy through the recollections of study participants and research collaborators. The result is an enlightening read, and a ripping yarn.

All three studies featured a phase in which the two groups competed for scarce resources such as prized penknives. In other respects, their designs were quite different. In the 1949 and 1953 studies, the boys underwent a phase of making friends. They were then assigned to one of two distinct groups that cut across friendship lines.



Move Fast and Break Things

Jonathan Taplin PAN MACMILLAN (2018)

With Facebook, Google and Amazon monopolizing consumer culture, digital-media expert Jonathan Taplin argues that their dominance is an economic war as well as a cultural one. His solution? A “digital renaissance” returning to principles of decentralization.

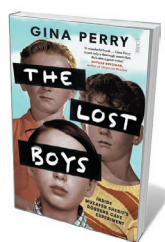


Resurrecting the Shark

Susan Ewing PEGASUS (2018)

Helicoprion, a bizarre prehistoric shark with teeth set in a spiral whorl, swam the oceans more than 270 million years ago. It remains shrouded in mystery. Susan Ewing traces how the fossil obsessed scientists for centuries, and how new research could resolve how its teeth fit into its jaw.

In the 1954 study, at Robbers Cave State Park in Oklahoma, there was no initial friendship phase. Moreover, competition was followed by a period in which the two groups could achieve a prized outcome (such as watching a movie) only if they cooperated (say, by pooling group funds). The studies were very much of their time: the scientists selected white, Protestant boys who were deemed psychologically ‘well adjusted’.



The Lost Boys: Inside Muzafer Sherif's Robbers Cave Experiments
GINA PERRY
Scribe (2018)

As Sherif and his colleagues reported in later texts — notably the 1966 book *Group Conflict and Co-operation* — their manipulations profoundly affected the boys' behaviour. In particular, as predicted by ‘realistic conflict’ theory, competition generally led to ‘us–them’ group identities: well-mannered boys were turned into aggressive, prejudiced adversaries. Significantly, at Robbers Cave, this process was then reversed with the requirement to cooperate in the study's final phase.

Sherif's research is less well known than Milgram's, or later classic studies by Solomon Asch on conformity and Philip Zimbardo on tyrannical power dynamics (B. Maher *Nature* **523**, 408–409; 2015). But what has made Sherif's legacy clearer and more enduring is the meticulous theoretical work that informed his studies' design. Sherif was no blind experimentalist. Rather, his ambitious goal was to create an empirical landscape capable of capturing the richness of ‘big picture’ social relations.

In many ways, this concern was a reflection of his own tumultuous life. As Perry clearly documents, that had been marked by external conflicts and inner torture. Before and after the Second World War, Sherif had moved back and forth between his native Turkey and the United States in the face of threats posed by nationalism, Nazism and McCarthyism. At various points, these pressures placed his work — sometimes his life — under threat, and led him to win and lose many friends along the way.

The Lost Boys illuminates Sherif's life and times, as well as Turkish history and

how large field studies work. Sherif's own accounts of the latter give a sense that support for his theoretical hypotheses followed reasonably seamlessly from the studies' manipulations. In practice, it wasn't quite like that, as Perry's careful detective work reveals.

First, the boys responded in a range of ways to changing group relations and escalating conflict, and it is not always easy to weave these into a single account. Second, even when they were describing the same event, Sherif's co-investigators often interpreted it differently. Third, it was impossible for the investigators not to shape the boys' behaviour — not least because ‘doing nothing’ was itself laden with significance

WELL-MANNERED BOYS WERE TURNED INTO AGGRESSIVE, PREJUDICED ADVERSARIES.

(as when researchers refused to censure intergroup aggression, and the tacit approval led to escalation). Fourth, sometimes things simply didn't go to plan. This is seen most vividly in the 1953 study, which — to Sherif's dismay — had to be abandoned because the boys, realizing the tensions were engineered, refused to buy into group conflict.

Perry does a magnificent job of documenting these nuances. She tracks down participants, many now retired, and shares their reactions on first discovering that they had taken part in a famous study. Most were intrigued and hungry for information; some were conflicted. Perry rightly worries about the ethics of her own psychological archaeology.

Nevertheless, her efforts to fill in the inevitable gaps in her sources are not always convincing. Sometimes she does rather too much ‘imagining’ to join the dots between experimenters' actions and participants' reactions. This is especially problematic in the context of her rather unforgiving

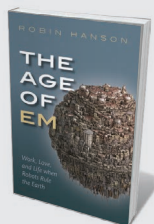
commentary on similar shortcomings in accounts by Sherif and his team. Although she questions whether Sherif's data collection was merely fleshing out a preconceived script, she herself is not immune to this charge.

A bigger problem is that Perry does not put the material she excavates to better use. Had she more thoroughly surveyed contemporary social psychological research on group conflict and collaboration, she would have found important clues that fit closely with the evidence she unearthed, and pave the way for significant progress in the questions that Sherif posed.

For example, in his 1976 monograph *Social Psychology and Intergroup Relations*, Michael Billig observed that Sherif's key theoretical failing was not factoring in the experimenters as the studies' third group. Michael Platow and John Hunter have pointed out that Sherif himself recognized that the effects of group membership (such as in-group affinity) preceded competition, and so seem to be as dependent on internalized group identity as on the battle for scarce resources (in ways that Henri Tajfel and John Turner would later unpack in their social identity theory). More generally, Sherif failed to appreciate how the participants and researchers would follow his own lead (in particular, in his cultivation of shared identity). As research has since clarified, this is a blind spot in many classic social psychology studies — not least those of Milgram and Zimbardo.

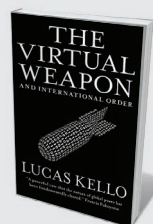
In *The Lost Boys*, Perry opens the door to clearer theorizing about these crucial processes of identity and influence, but she fails to walk through it. In these terms, her book leaves the reader concerned not just for the boys' lost voices, but for Sherif's. He argued passionately and compellingly for theoretical progress in social psychology. Today, when a focus on empirical replicability often drowns out the equally important requirement for strong integrative theory, we need that voice as much as we did 70 years ago. ■

Alex Haslam is professor of psychology and Australian Laureate Fellow at the University of Queensland in Brisbane. His most recent book is *The New Psychology of Health* (with Catherine Haslam, Jolanda Jetten, Tegan Cruwys and Genvieve Dingle). e-mail: a.haslam@uq.edu.au



The Age of Em: Work, Love, and Life when Robots Rule the Earth

Robin Hanson OXFORD UNIV. PRESS (2018)
Marshalling economics, physics and philosophy, Robin Hanson predicts a future run by brain emulations (“ems”), featuring era-specific issues such as “mind theft”. Hanson's predictions detail a world both uncanny and eerily familiar.



The Virtual Weapon and International Order

Lucas Kello YALE UNIV. PRESS (2018)
The cyber revolution clearly constitutes an ever-growing challenge to international order. Lucas Kello reflects on technology's role in political revolution, and the importance of aligning international-relations studies with the unruly expansion of cyberspace.

GENETICS

Pinpointing the pain gene

Tor Wager lauds a book on the hunt for an elusive root of sensory suffering.

Two soldiers receive similar injuries in battle. One recovers in months; the other endures excruciating pain for years. Why this difference?

The question is pressing. One in five people suffers from chronic pain, affecting every aspect of their lives. Although significant gains have been made with anaesthetics and anti-inflammatory medications, the roots and relief of long-term pain are proving harder to find. Pain is also fuelling a global epidemic of opioid addiction and related deaths. In *Chasing Men on Fire*, neurologist Stephen Waxman provides a compelling portrait of scientific discovery in this area.

Waxman, who works in basic research and clinical medicine, offers an insider's account of the global search for a pain gene, beginning in 1966. He intertwines descriptions of cross-disciplinary neuroscience with portraits of scientists, and the struggles of people with conditions such as erythromelalgia, or 'man-on-fire syndrome', characterized by burning pain in hands and feet. Structurally, the book is innovative: 11 research papers are interlaced with the stories behind them. It is thus both a boon for researchers and an engrossing read for nonspecialists.

Humans love simplicity. We want the intricate systems in our brains and bodies to 'just work'. But Waxman shows that biology is complex, and genetic clues can be elusive. Detecting them relies on finding regularities across many people, which can make it seem impossible to pinpoint a key gene, and the rare mutations in it that lead to disease. As he reveals, it took considerable sifting and coordinated effort on three continents by scientists including pharmacologists, electrophysiologists, molecular biologists and geneticists before a 'master gene' for pain was isolated.

That gene, *SCN9A*, encodes the complex molecule $\text{Na}_v1.7$, a sodium channel and a basic building block of nervous-system function. When activated by electrical

current (for example, from stimulation of nerve endings), sodium channels allow ions to rush into neurons, causing the cell to fire a nerve impulse. $\text{Na}_v1.7$ channels are found nearly exclusively in peripheral nociceptive neurons in the dorsal root ganglia; these nerve centres in the spinal column are the first 'relay station' in transmitting pain-related information to the brain. This makes the channels potential targets for therapies that alleviate pain without altering brain function catastrophically.

Over more than 20 years, Waxman worked with dozens of people with erythromelalgia. His team painstakingly sequenced each person's *SCN9A* gene, hunting for mutations. As he puts it, he was "navigating a large, complex sea". To understand the functional implications of the mutations, he and his team extracted adult blood cells and turned them back into stem cells. They then triggered these cells to grow into $\text{Na}_v1.7$ -expressing neurons in the dor-

"People are suffering, and we have the responsibility to help them if we can."

sal root ganglia, to study how each individual's unique mutation affected the cells' properties. They discovered dozens of rare pain-causing mutations. Later, they used computer models to understand how the mutations affect the sodium channels' structure, allowing them to predict whether a person would respond to a particular drug.

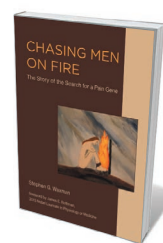
This research has already helped some people with chronic pain by providing an explanation for its cause, and identifying drugs that are effective for some. Compounds that interact with $\text{Na}_v1.7$ channels to alleviate pain are currently in clinical trials. Understanding the mechanisms can help the rest of us: genetic variation in the channels affects susceptibility to common sources of chronic pain, such as surgery. (This does not

mean that molecular mechanisms are the only important ones. Post-surgical pain can be strongly affected by emotions such as anxiety, and the complex brain circuitry underlying them.)

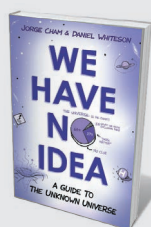
Embedded in Waxman's narrative are broader lessons. First, we need to face complexity head-on. One gene can go wrong in thousands of ways. Each of the 1,800 amino acids in *SCN9A* presumably affects how the protein it encodes folds. That, in turn, determines whether the channel opens and closes properly to transmit pain-related signals at appropriate times. The story of *SCN9A* reveals how the pursuit of basic understanding lays a crucial foundation for clinical advances once undreamt of. With the pain gene, this pursuit stretches from scientific experiments on squid by Alan Hodgkin and Andrew Huxley in the mid-twentieth century to the reconstruction of ion channels' crystal structures by Waxman's group, and beyond.

Waxman's story is also deeply human. It pivots on cross-border, cross-disciplinary scientific collaboration in service of the greater good. It demonstrates a pursuit of scientific understanding that keeps sight of the big picture: that people are suffering, and we have the responsibility to help them if we can. Finally, it conveys the spirit of how science at its best is accomplished — with urgency, passion, inventiveness and collaboration. In Waxman's words: "Let's just do it." ■

Tor Wager is a professor of psychology, neuroscience and cognitive science at the University of Colorado Boulder, and director of the Cognitive and Affective Neuroscience Lab. e-mail: tor.wager@colorado.edu



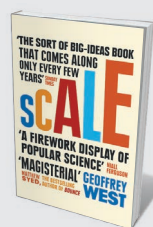
Chasing Men on Fire: The Story of the Search for a Pain Gene
STEPHEN G. WAXMAN
MIT Press (2018)



We Have No Idea: A Guide to the Unknown Universe

Jorge Cham and Daniel Whiteson RIVERHEAD (2018)

This cheerily conversational exploration of grey areas and conundrums, from the composition of the cosmos to the elements, is peppered with cartoons. Jorge Cham and Daniel Whiteson are upbeat guides to universal ignorance.



Scale

Geoffrey West WEIDENFELD & NICOLSON (2018)

In this "grand unified theory of sustainability", physicist Geoffrey West explores underlying laws that link society and nature, called scaling theory. Insights (into city size and walking speed, for instance) abound. (See P. Ball *Nature* **545**, 154–155; 2017.) [Mia Gray](#)

Correspondence

Tracking ecology is not old-fashioned

Long-term monitoring is not innovative and is never going to result in spin-off businesses, but it is still the best way of observing human impact on the environment. The use of trained personnel to take outdoor measurements is being called into question, however.

The UK Countryside Survey, for example, relies on experienced botanists to go out in all weathers to find out how habitats and species are changing. Their ability to recognize a particular species of grass among other vegetation, for instance, cannot be replaced by technology. In my view, expert volunteers (citizen scientists) who are willing to monitor particular ecological environments or species are no substitute, because their botanical expertise does not usually focus on the commonplace and the widespread.

Doing the same observations in the same way and at the same places no longer seems to light up potential funders. If they underrate the valuable expertise of fieldworkers and field botanists, we stand to lose one of the most highly regarded ecological monitoring programmes in Europe and the world.

Lisa Norton *Centre for Ecology and Hydrology, Lancaster, UK.*
l.n@ceh.ac.uk

Code review poses extra challenges

As editors-in-chief of *ReScience*, a journal dedicated to computational replication (see <http://rescience.github.io>), we argue that the review of software codes used in research papers requires a departure from the conventional refereeing process (see *Nature* 555, 142; 2018).

In addition to scientific expertise, reviewers need experience in software development and programming languages to be able to run

the code and inspect it. They must test the source code and whether figures and/or tables from the submitted article can be reproduced using the software, input files, data sets and instructions supplied by the authors.

In our experience, technical problems that arise in installing and running scientific software are resolved most effectively if authors and reviewers can discuss the issues with one another. *ReScience* uses the GitHub platform for such open reviewing (see also N. P. Rougier *et al.* *PeerJ Comp. Sci.* 3, e142; 2017).

Nicolas P. Rougier *Inria Bordeaux — South West Research Centre, Talence, France.*
Konrad Hinsén *Centre for Molecular Biophysics (CBM), CNRS, Orléans, France.*
konrad.hinsén@cncrs.fr

Quality-assured data for enzyme activity

Transparent reporting of experimental methods to ensure the reproducibility of results (see *Nature* 555, 6; 2018) is particularly crucial in enzyme kinetics, given the wide variation in many assay parameters. It also allows unambiguous interpretation of the data.

The STRENDA (Standards for Reporting Enzymology Data) database is a repository for published enzymology data, lodged under its guidelines for transparent reporting of experimental methods (see go.nature.com/2qq9vm7). The STRENDA Commission is working with the biochemistry community and funding agencies to make submission to this database routine practice during publication. More than 50 journals currently recommend the guidelines to their authors.

The implementation of scientific data standards has typically been left to data curators (see, for example, U. Wittig *et al.* *Nucl. Acids Res.* 46, D656–D660; 2018). The STRENDA database now serves as a formal validation

tool for reliable reporting of data (see also N. Swainston *et al.* *FEBS J.* <https://doi.org/cm8d>; 2018).

We suggest that our model could be adapted for the benefit of authors, reviewers, data consumers, publishers and funders across experimental disciplines.

Neil Swainston, Carsten Kettner *on behalf of the STRENDA Commission, Beilstein Institute for the Advancement of Chemical Sciences, Frankfurt am Main, Germany.*
ckettner@beilstein-institut.de

A surge in Brazilian papers in top journals

As a rough assessment of Brazil's contribution to high-impact science from 1980 onwards, we analysed the number of papers published in *Nature* and *Science* from three of the country's leading universities. We found a dramatic increase in their publications in these prestigious journals over the period.

We combined publication counts for the University of São Paulo, the University of Campinas and the Federal University of Rio de Janeiro. To track the long-term trend in their performance, we sampled counts every decade from 1980 to 2010 and for 2017. During the 5 individual years we sampled from this span of 37 years, these institutions together published 0.08 papers, on average, in each edition of the two journals (details available from authors on request).

Although this is low compared with averages amounting to about 1.5, 0.6 and 0.5 papers per edition from Harvard University in Cambridge, Massachusetts, and the UK universities of Cambridge and Oxford, respectively, the total number of papers from the Brazilian universities increased by 2,200% from 1980 (1 article) to 2017 (23 articles). The rise has been steepest during the current decade (from just 7 papers in 2010), despite Brazil's economic

crisis that started in 2014.

This overall increase, in our view, reflects the intense research activity and resilience of the Brazilian institutions, which all have a strong history of international collaboration. Now more than ever, governmental commitment to science is crucial for our future research performance.

Gabriel José de Carli, Tiago Campos Pereira *University of São Paulo, Brazil.*
tiagocampospereira@ffclrp.usp.br

Set goals for cancer research funding

Cancer Research UK — one of the country's largest voluntary-sector funding organizations — is seeking to improve the survival rates of people with cancer from around 50% today to 75% in 2034. We propose ways to focus cancer-research funding more effectively to help attain this goal.

We suggest that priority funding should go towards improving detection of early-stage, treatable tumours; developing innovative therapies for cancers that have high mortality rates even when they are detected early; and accelerating the translation of promising new drugs — which currently takes 17 years on average (Z. S. Morris *et al.* *J. R. Soc. Med.* 104, 510–520; 2011).

Strong leadership is crucial for implementing this three-pronged approach, and for optimal oversight of research funding. From our experience, training of team leaders needs to concentrate more on developing the skills for managing research teams and on acquiring the techniques for efficiently directing processes. Such training could be delivered by a leadership academy run by respected research leaders and business coaches.

Johnathan Watkins, Wahyu Wulaningsih *PILAR Research Network, Cambridge, UK.*
jwatkins@pilar.org.uk

Immune memory in the brain

The brain's resident immune cells retain a long-lasting memory of peripheral inflammation. This memory can influence the response to stroke and the progression of Alzheimer's disease in mouse models. [SEE ARTICLE P.332](#)

ALEXI NOTT & CHRISTOPHER K. GLASS

The body's innate immune system provides an immediate and nonspecific defence against invading pathogens. The main innate immune cells in the brain, microglia, rarely encounter such infections, but they can respond to peripheral inflammation elsewhere in the body. One intriguing feature of peripheral innate immunity is a phenomenon called immune memory: the body's innate immune system 'remembers' previous exposures to microbes, and augments its responses to reinfections accordingly^{1,2}. Does it follow, then, that peripheral inflammation could elicit immune memory in microglia? On page 332, Wendeln *et al.*³ provide evidence for immune memory in microglia, and show that it can alter the progression of brain disorders in mice.

There are two types of immune memory: training, which exaggerates immune responses; and tolerance, which dampens them. To test whether microglia retain an immune memory of peripheral infection, Wendeln *et al.* injected lipopolysaccharide (LPS) molecules, a component of some bacteria, into the body cavities of mice. Exposing wild-type mice to two injections of LPS induced a microglial response resembling immune training, characterized by elevated levels of pro-inflammatory molecules. Four LPS exposures resulted in immune tolerance, indicated by reduced pro-inflammatory signals.

Peripheral immune training improves the immune system's ability to eliminate reinfections, but it can be harmful in people who have inflammatory conditions. By contrast, peripheral immune tolerance can be undesirable for clearing reinfections, but it is beneficial in organs that are subject to continuous pathogen exposure, such as the gut². Changes in innate immunity in the brain are implicated in a diverse spectrum of disorders⁴, but the contribution of immune memory to these has not been explored. Wendeln *et al.* therefore investigated whether immune training or tolerance in microglia could influence the progression of Alzheimer's disease and stroke, using mouse models.

A hallmark of Alzheimer's disease is the build-up of amyloid- β protein, which activates

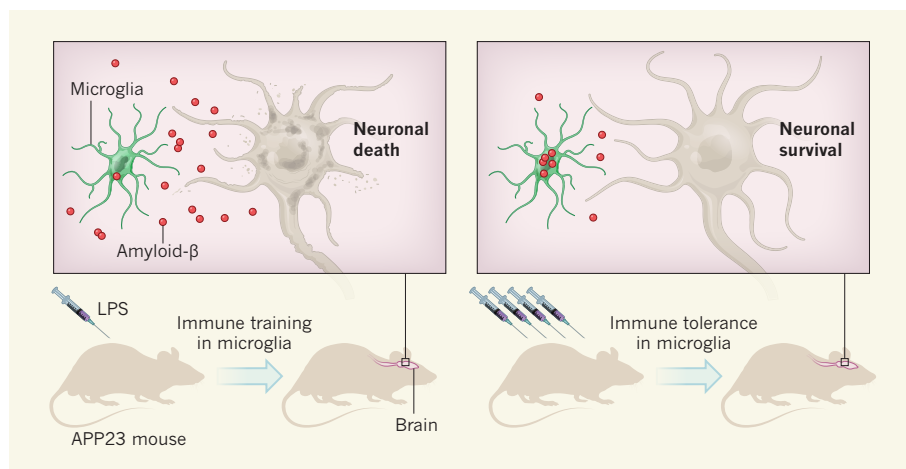


Figure 1 | Immune memory in a model of Alzheimer's disease. Innate immune cells can hold a 'memory' of previous exposure to pathogens that alters future immune responses. Wendeln *et al.*³ investigated innate immune memory in a mouse model of Alzheimer's disease (a strain dubbed APP23), in which the protein amyloid- β accumulates in the brain. The authors injected lipopolysaccharide (LPS) molecules into the animals' body cavities either once or four times, to mimic bacterial infection. In mice that were injected once, brain immune cells called microglia exhibit a form of immune memory called immune training, which exaggerates inflammatory responses. After six months, amyloid- β levels are higher than those in uninjected mice, and more neurons die. By contrast, the microglia of mice injected four times exhibit immune tolerance, which dampens inflammatory responses. This treatment leads to an increase in microglial amyloid- β uptake and better neuronal survival.

microglia⁵. The authors used a mouse model of Alzheimer's disease (a strain dubbed APP23) that recapitulates amyloid- β accumulation in the brain. In this model, a single dose of LPS could induce microglial training, probably because the presence of amyloid- β provided an additional pro-inflammatory stimulus. Four doses of LPS induced a tolerant response (Fig. 1). The researchers analysed the animals' brains after six months.

Wendeln and colleagues found that immune training increased amyloid- β accumulation and that immune tolerance decreased amyloid- β , compared to levels in untreated animals of this strain. Similarly, immune tolerance reduced levels of neuronal death seven days after the researchers induced stroke in wild-type mice.

Wendeln *et al.* found that microglial immune memory persisted in APP23 mice for at least six months. Peripheral immune memory seems to be wired into stem cells, and so can be propagated in the cells' progeny for long time periods^{6–8}. However, there is no evidence for the existence of microglial stem

cells. Instead, immune memory in microglia might endure because these cells are long-lived⁹. And persistent brain inflammation, initiated by amyloid- β accumulation, could help to maintain immune memory.

What mechanisms might sustain microglial immune memory? Regions of DNA called enhancers can boost the expression of nearby genes when activated, thereby controlling cell behaviour. Enhancers in bone-marrow-derived immune cells can become activated or poised for activation following transient peripheral inflammation and can persist in this state after inflammation ceases, providing a 'memory' of past events¹⁰. Wendeln *et al.* showed that treating APP23 mice with LPS increased levels of enhancer activation in microglia, compared to treating healthy mice with LPS. This indicates that long-lasting enhancer activation in microglia requires both brain pathology and an inflammatory insult.

The researchers demonstrated that microglial tolerance led to enhancer and gene-expression changes associated with a process called phagocytosis, by which microglia

ingest material to be broken down. These changes correlated with increased microglial uptake of amyloid- β . By contrast, training led to enhancer and gene-expression changes associated with inflammation and energy expenditure.

Could targeting enhancer activity be a way of regulating immune memory in the brain, thereby altering the progress of neurological disorders? Enhancer activation is modulated by histone deacetylase (HDAC) enzymes, and the authors found that the loss of HDAC2 in microglia blocked immune training in these cells in healthy mice. Reduction of HDAC2 levels in neurons has been shown to improve memory and reverse gene-expression changes induced by neurodegeneration¹¹. Furthermore, loss of HDAC1 and HDAC2 in microglia reduces amyloid- β levels and improves memory in a different mouse model of Alzheimer's disease¹². An appealing possibility is that HDAC2 inhibition blocks microglial immune training and enhances the cells' ability to clear amyloid- β in people with Alzheimer's disease. However, regulating immune memory in the brain without causing deleterious consequences in the rest of the body will be challenging.

This work also opens up other avenues for research. For instance, metabolic products are often required for the activity of enzymes that regulate enhancers. Perhaps changes in metabolism mediate and perpetuate enhancer activation in microglia. Whether this is the case, and how specific enhancers are then targeted for training or tolerance, remains unknown.

The mechanism by which immune memory is transmitted to the brain from the periphery is also unknown. One possibility, suggested by Wendeln *et al.*, is that inflammatory molecules are transported to the brain through the blood. Alternatively, peripheral immune cells might infiltrate the brain and activate microglia, or peripheral nerves might send signals to indicate that inflammation has occurred. One such possible neuronal pathway is the gut-brain axis, through which gut microbes can modulate microglial behaviour¹³.

The authors focused on microglia, but other cell types in the brain probably also contribute to the effects of training and tolerance. For example, cells called astrocytes, which have immunomodulatory functions, are activated during Alzheimer's disease¹⁴. The researchers showed that there were fewer activated astrocytes in APP23 mice exposed to either one or four doses of LPS than there were in untreated APP23 mice. This difference might be driven by microglia-derived factors¹⁵ — but another explanation is that astrocytes, which can detect inflammatory signals¹⁴, also retain a memory of previous stimuli. These possibilities still need to be tested.

Although rapid progress has been made in identifying genetic contributions to neurological disorders, environmental factors also have substantial effects. Because of this, Wendeln and co-workers' findings will be of broad interest. In

addition to changes due to infections, the innate immune system can be influenced by environmental stimuli such as stress. Stress has been linked to neurological disorders, and an effect of stress on innate immune training has been hypothesized but remains unexplored in the context of these diseases¹⁶.

Finally, blocking immune training or mimicking immune tolerance in the brain will be of therapeutic benefit only if the current findings are replicated in humans. Demonstrating microglial immune memory in humans will be difficult, owing to the inaccessibility of brain tissue in living people. However, analysis of inflammatory signalling molecules released into the cerebrospinal fluid could be used as a proxy for microglial immune memory. Regardless of the immediate therapeutic potential, Wendeln and colleagues' work sets the stage for further investigation of the impact of environmental factors on microglial function in neurodegenerative conditions. ■

Alexi Nott and Christopher K. Glass are in the Department of Cellular and Molecular

Medicine, University of California, San Diego, San Diego, California 92093, USA. C.K.G. is also in the Department of Medicine, University of California, San Diego.
e-mail: ckg@ucsd.edu

1. Netea, M. G., Latz, E., Mills, K. H. G. & O'Neill, L. A. J. *Nature Immunol.* **16**, 675–679 (2015).
2. Netea, M. G. *et al.* *Science* **352**, aaf1098 (2016).
3. Wendeln, A.-C. *et al.* *Nature* **556**, 332–338 (2018).
4. Labzin, L. I., Heneka, M. T. & Latz, E. *Annu. Rev. Med.* **69**, 437–449 (2018).
5. Sturchler-Pierrat, C. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 13287–13292 (1997).
6. Kaufmann, E. *et al.* *Cell* **172**, 176–190 (2018).
7. Mitroulis, I. *et al.* *Cell* **172**, 147–161 (2018).
8. Christ, A. *et al.* *Cell* **172**, 162–175 (2018).
9. Füger, P. *et al.* *Nature Neurosci.* **20**, 1371–1376 (2017).
10. Ostuni, R. *et al.* *Cell* **152**, 157–171 (2013).
11. Gräff, J. *et al.* *Nature* **483**, 222–226 (2012).
12. Datta, M. *et al.* *Immunity* **48**, 514–529 (2018).
13. Erny, D. *et al.* *Nature Neurosci.* **18**, 965–977 (2015).
14. Sofroniew, M. V. *Nature Rev. Neurosci.* **16**, 249–263 (2015).
15. Liddel, S. A. *et al.* *Nature* **541**, 481–487 (2017).
16. Salam, A. P., Borsini, A. & Zunszain, P. A. *Mol. Psychiatry* **23**, 170–176 (2018).

This article was published online on 11 April 2018.

NANOSCIENCE

A peptide-guided twist of light

The growth of gold nanoparticles has been manipulated using amino acids and peptides to produce twisted structures that alter the rotation of light. The method could simplify the development of optical devices. SEE LETTER P.360

GUILLERMO GONZÁLEZ-RUBIO & LUIS M. LIZ-MARZÁN

Nanoparticles that control the rotation of light have potential applications, for example in optical devices¹ and sensors², but preparing such particles has been difficult, especially from crystalline metals. On page 360, Lee *et al.*³ report a remarkable method that uses amino acids or peptides (small molecules formed from amino acids) to direct the dissymmetric growth of gold nanoparticles that have a twisted morphology. The findings open up radical opportunities for the preparation of materials and devices that control light rotation.

Dissymmetric objects that cannot be superimposed on their mirror image are found at a variety of scales and include molecules of DNA, snail shells and even galaxies. Such structures are said to be chiral. Louis Pasteur coined the concept of molecular dissymmetry in 1848, when he attributed the morphological differences in crystals of tartrate to the existence of mirror-image tartrate molecules^{4,5}. We now know that the functions of biomolecules

often depend on chirality, which, for example, provides the basis of exquisitely specific interactions between enzymes and their substrate molecules, enabling the proper functioning of living organisms.

One property of chiral molecules is that each mirror-image form interacts differently with circularly polarized light (in which the electric field traces a helix in the direction of the light's propagation), resulting in phenomena known collectively as optical activity. For example, circular dichroism involves the differential absorption of left- and right-handed circularly polarized light by the mirror-image forms of a molecule. The optical activity of chiral organic molecules has been used to manipulate the rotation of light, but almost invariably in the ultraviolet region of the electromagnetic spectrum.

In the past decade, some inorganic materials have also been shown to have chirality and optical activity⁶, thereby enabling control of the rotary propagation of light to be extended to the visible and near-infrared regions. Prominent among these inorganic compounds are nanostructured materials that

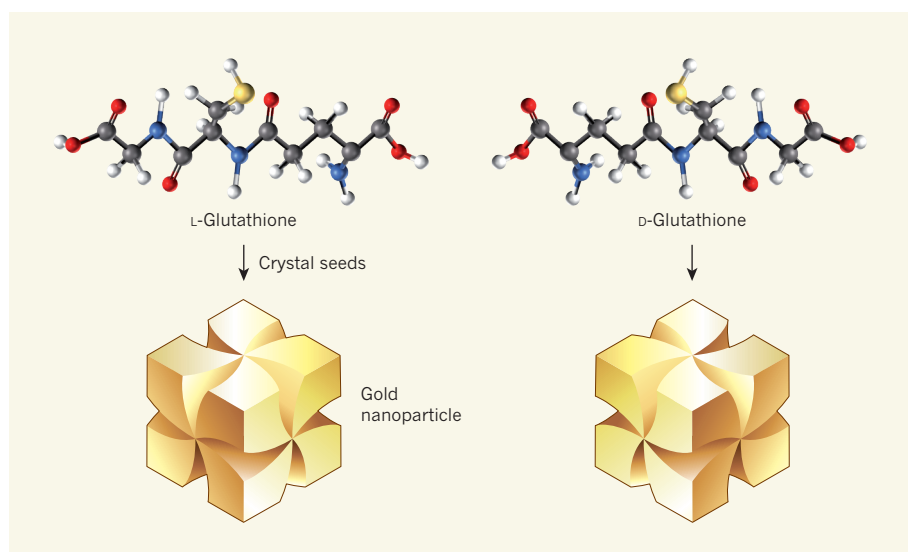


Figure 1 | The transfer of chirality from peptides to nanoparticles. Lee *et al.*³ grew gold nanoparticles from crystal ‘seeds’ in the presence of chiral amino acids or peptides, which can exist as mirror-image forms. The resulting nanoparticles were also chiral, and the mirror-image form that grew depended on the form of the amino-acid or peptide additive that was used. For example, the peptide glutathione can occur as mirror-image L- and D-isomers, which direct the growth of mirror-image versions of the helicoid-shaped nanoparticle shown. (Glutathione structures from Guillermo González-Rubio.)

exhibit plasmonic effects. Such effects derive from the oscillations of conduction electrons in nanostructured metals or in other materials that contain free electrons, and result in the extremely efficient absorption and scattering of visible and near-infrared light. The wavelength involved is defined by the composition, dimensions and morphology of the nanomaterial.

The use of chiral plasmonic effects has been identified as one of the most promising routes to developing optical metamaterials — artificial structures such as ‘invisibility cloaks’⁷ with optical properties that differ from those of materials found in nature. This has motivated considerable effort towards the fabrication of nanoscale objects that have chiral geometry. Substantial advances have been achieved¹ both through top-down fabrication methods, in which nanoscale objects are prepared from bulk materials, and through bottom-up methods, in which the objects are grown using chemical processes.

Top-down approaches can already be used to make small quantities of nanomaterials that have well defined morphologies, but it might be difficult to scale up these approaches to produce the amounts that will be needed for processing into materials or integration into devices. By contrast, bottom-up approaches are typically based on chemistry performed in solution, which is easier to scale up.

Remarkable advances have been made in the scaling up of bottom-up methods to make chiral nanomaterials, mainly by using a chiral template to direct the assembly of preformed nanoparticles. Beautiful examples of such materials include gold spheres adsorbed onto DNA strands⁸, gold nanorods interleaved with accurately programmed DNA-origami

structures⁹ and gold nanorods adsorbed onto helical protein fibres². But in all of these cases, the optical activity obtained is the result of collective plasmonic effects, and the wavelength at which circular dichroism occurs is defined both by the specific properties of the individual building blocks used and by their organization on the template. This means that several parameters must be manipulated to achieve a specific optical effect.

A simpler alternative for generating optical effects would be to grow chiral plasmonic nanoparticles in a way that ensures that all such particles have the same morphology

“The nanomaterials could be processed into composite materials, and might find technological applications.”

and, therefore, identical optical activities. This can be achieved by using preformed nanoparticles as ‘seeds’, which are then grown to the desired size and shape by the slow precipitation of material onto them, typically using additive molecules to direct the growth process¹⁰. Such methods have been used to make highly symmetric metallic nanoscale objects, including spheres, rods and octahedra. This approach has also been used to make chiral structures from certain inorganic materials¹¹, but not from metals such as gold that have a highly symmetric crystalline structure. Lee *et al.* now report an advance that fills this methodological gap.

The main breakthrough that the authors report is the use of chiral amino acids or peptides that contain thiol (SH) groups as additives in the seeded growth of gold nanoparticles (Fig. 1). These additives affect the

growth rate of certain crystal facets, which leads to the formation of nanostructures that have intricate chiral morphologies and an impressive degree of monodispersity — all particles are highly similar in size and shape. Moreover, the obtained morphology can be manipulated by varying either the structure of the shape-directing molecule or the initial shape of the seed particles.

Lee and colleagues therefore demonstrate that the chirality and optical behaviour of naturally occurring amino acids and peptides can be transferred to shaped plasmonic nanocrystals. The resulting high-quality, chiral gold nanoparticles (see scanning electron microscopy images in Fig. 1 of ref. 3) show strong circular dichroism (a large difference between the absorption of left- and right-handed circularly polarized light), with the wavelength and intensity of the signal determined by the nanoparticles’ specific morphology. Because this remarkable optical response arises from intrinsic single-particle effects, the nanomaterials could be processed into composite materials or thin films, and might even find technological applications through incorporation into devices.

The authors’ procedure is a remarkably simple modification of methods that are commonly used to grow shaped gold, silver or palladium nanoparticles. It is therefore likely to be readily adopted to produce chiral nanostructures of these ‘noble’ metals, which have improved catalytic or electronic properties compared with analogous non-chiral structures. The success of the technique will depend on whether it does indeed work for noble metals other than gold, and whether the small, naturally occurring chiral additives can be replaced by synthetic dissymmetric molecules. Further studies are needed to determine how the process is affected by the growth kinetics of particles, by the strength of the interactions between the nanocrystal surface and the chiral additive, and by the composition and size of the seeds. ■

Guillermo González-Rubio and Luis M. Liz-Marzán are at CIC biomaGUNE, 20014 Donostia-San Sebastián, Spain. L.M.L.-M. is also at Ikerbasque, Basque Foundation for Science, Bilbao, Spain. e-mail: llizmarzan@cicbiomagune.es

- Hentschel, M., Schäferling, M., Duan, X., Giessen, H. & Liu, N. *Sci. Adv.* **3**, e1602735 (2017).
- Kumar, J. *et al. Proc. Natl Acad. Sci. USA* **115**, 3225–3230 (2018).
- Lee, H.-E. *et al. Nature* **556**, 360–365 (2018).
- Pasteur, L. *Ann. Chim. Phys. Sér. 3* **24**, 442–459 (1848).
- Gal, J. *Nature Chem.* **9**, 604–605 (2017).
- Ma, W. *et al. Chem. Rev.* **117**, 8041–8093 (2017).
- Pendry, J. B. *Science* **306**, 1353–1355 (2004).
- Kuzzyk, A. *et al. Nature* **483**, 311–314 (2012).
- Lan, X. *et al. J. Am. Chem. Soc.* **137**, 457–462 (2015).
- Liz-Marzán, L. M. & Grzelczak, M. *Science* **356**, 1120–1121 (2017).
- Zhou, Y., Yang, M., Sun, K., Tang, Z. & Kotov, N. A. *J. Am. Chem. Soc.* **132**, 6006–6013 (2010).

EVOLUTION

When sex differences lead to extinction

There are striking differences between the male and female forms of some species. A study of marine fossils finds that such differences come at the cost of an increased risk of extinction. [SEE LETTER P.366](#)

HANNA KOKKO

It is tempting to imagine evolution as a process of endless improvement. Sexually reproducing organisms continually shuffle their genomes, trying out ever more tricks that might help the organisms to cope with the challenges they face. So why does extinction sometimes occur? Is it sheer bad luck, or can selection give rise to traits that cause a species to enter a 'danger zone' of heightened risk of extinction? On page 366, Martins *et al.*¹ report their examination of the fossil record, and identify one such danger zone. When the males and females of a given species look substantially different, this correlates with the fossil-record presence of such species having a shorter time span than that of species in which males and females look similar.

Martins and colleagues studied fossilized ostracods, which are crustacean species of the class Ostracoda. These tiny aquatic animals look like a cross between a shrimp and a mussel. They are evolutionarily close to the former. Ostracods make up a much greater part of the fossil record than their more-famous distant cousins, the trilobites. This is good news if ostracod samples are being used to calculate estimates of extinction risk.

Moreover, thousands of ostracod species exist today. Therefore, it is known that the shape of the ostracod exoskeleton (the fossilizable part of the animal) can be used to distinguish males from females. Males are more elongated than females, because they need extra space for their reproductive organs (see Fig. 1 of ref. 1). However, in terms of overall body size, in some species the males are larger than the females, whereas in others the females are the larger sex. The scale of the differences in size and shape between males and females can range from being relatively small to being highly pronounced, depending on the species.

The authors analysed the shapes and sizes of fossil exoskeletons of 93 ostracod species. These ostracods inhabited what is now eastern Mississippi between 84 million and 66 million years ago, during the Late Cretaceous period, at a time when an interior sea split North America into eastern and western halves. The authors' analysis of fossils, along with a statistical modelling

approach, enabled them to uncover a curious pattern. When comparing species, it emerged that those in which males were very different from females had a poorer prognosis for continued existence. The authors' models predict a tenfold increase in extinction risk per unit time when species in which males are larger than females, with large differences in shape between the sexes, are compared with species in which the males are smaller than the females, with small differences in shape between the sexes.

The importance of this finding for our understanding of evolution makes it of interest to more than just ostracod enthusiasts.

Sexual reproduction opens the door for sexual selection, the selection of characteristics that promote successful mating. Therefore, the generation of offspring requires both survival skills and the ability to compete for opportunities to reproduce. This can drive different selection pressures for males and females, and there is a growing appreciation in evolutionary biology that sex differences have the potential to either help or hinder the persistence of entire populations or species.

If males invest heavily in characteristics that aid different tasks from those undertaken by females, the population could benefit if strong selection weeds out suboptimally performing males and leads to the species' genome becoming better adapted over time². However, there is also a danger that selection for male reproductive success could result in characteristics that are harmful to females³, whose ability to reproduce is more valuable for population persistence than is male input. Indeed, males of the common lizard (*Lacerta vivipara*) compete so intensely that their aggressive behaviour and biting of females reduces female life span and population growth⁴. Another simpler and perhaps under-studied effect is that, from a population-growth perspective, large, growing males consume resources that could have



Figure 1 | The courtship dance of a male peacock spider (*Maratus speciosus*). Certain species have marked differences between the male and female forms — for example, male peacock spiders are strikingly more colourful than their female counterparts. Studying fossilized aquatic creatures called ostracods, Martins *et al.*¹ investigated whether the degree of difference between male and female forms of a given species affects its risk of extinction.

been put to better use if left for females⁵.

Why doesn't evolution favour asexual reproduction, to avoid the types of struggle between males and females that can have a negative effect on overall population fitness? Theoretically, an animal lineage that reproduces asexually should eventually run into difficulties. However, it was reported⁶ this year that Amazon molly fish (*Poecilia formosa*) have a genome of surprisingly good quality, even after about 500,000 generations of asexual reproduction. Many such evolutionary mysteries will provide fascinating research topics for years to come.

Martins and colleagues' approach overcomes caveats in previous attempts to measure how differences between the sexes affect population fitness. In a technique termed experimental evolution, the strength of sexual selection can be varied experimentally by restricting mating in some captive lineages to between monogamous pairs, while allowing competition for mates to operate in other lineages of the same species⁷. Another approach involves using information on species existing today and indicators of sexual competition, such as testis size or

differences in size or colour between males and females. These indicators are then compared with the estimated risk of species extinction as documented⁸ in the Red List generated by the International Union for Conservation of Nature, or with the levels of population turnover⁹. However, experimental evolution is usually carried out in a simplified laboratory environment, whereas current threats to species persistence often have human-mediated causes. Martins and colleagues managed to show the risks of pronounced male–female differences over a long period before humans had evolved.

We can thank sexual selection for wondrous traits such as the peacock's tail, the courtship dance (Fig. 1) of the colourful male peacock spiders of the genus *Maratus* and, indeed, the elongated shape of male ostracods. However, as Martins *et al.* have shown, differences between the sexes can have negative consequences for species. With more than 10,000 ostracod species still in existence (including asexual ones), it is surprising how little we know about their genetics or other demographic factors that affect how these populations thrive, including the conditions under which

they reproduce or survive well¹⁰. Why do differences between male and female ostracods result in an increase in the risk of extinction? Experimental evolution, anyone? ■

Hanna Kokko is in the Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland.

e-mail: hanna.kokko@ieu.uzh.ch

1. Martins, M. J. F., Puckett, T. M., Lockwood, R., Swaddle, J. P. & Hunt, G. *Nature* **556**, 366–369 (2018).
2. Whitlock, M. C. *Evolution* **54**, 1855–1861 (2000).
3. Mank, J. E. *Nature Rev. Genet.* **18**, 721–730 (2017).
4. Le Galliard, J.-F., Fitz, P. S., Ferrière, R. & Clobert, J. *Proc. Natl Acad. Sci. USA* **102**, 18231–18236 (2005).
5. Kokko, H. & Brooks, R. *Ann. Zool. Fennici* **40**, 207–219 (2003).
6. Warren, W. C. *et al. Nature Ecol. Evol.* **2**, 669–679 (2018).
7. Lumley, A. J. *et al. Nature* **522**, 470–473 (2015).
8. Morrow, E. H. & Fricke, C. *Proc. R. Soc. Lond. B* **271**, 2395–2401 (2004).
9. Doherty, P. F. *et al. Proc. Natl Acad. Sci. USA* **100**, 5858–5862 (2003).
10. Park, A. W., Vandekerhove, J. & Michalakakis, Y. J. *Evol. Biol.* **27**, 1650–1661 (2014).

This article was published online on 11 April 2018.

ENGINEERING

Electronics and photonics united

A method for integrating photonic devices with state-of-the-art nanoelectronics overcomes previous limitations. The approach shows promise for realizing high-speed, low-power optoelectronic technology. SEE LETTER P.349

GORAN Z. MASHANOVICH

The integration of electronic and photonic circuits on a single silicon chip could enable unprecedented functions and performance in computing, communications and sensing at a low cost. But this goal has been hindered by the fact that most electronic circuits use bulk silicon substrates, whereas photonic circuits typically require silicon-on-insulator platforms. On page 349, Atabaki *et al.*¹ report the first fabrication of photonic devices on a bulk silicon substrate, together with millions of electronic devices known as transistors. The work paves the way for the mass production of optoelectronic systems on chips.

Photonics is prevalent in almost every aspect of day-to-day life — from smartphones and display screens to lighting and medical devices. It is often considered to be the 'electronics of the twenty-first century'. Although silicon is not an ideal photonics material (for example, lasers cannot be built from silicon), many factors have made it the main candidate

for applications that require large numbers of photonic devices². These factors include the high natural abundance of silicon, its widespread use in electronics, its optical transparency over a wide range of wavelengths and the availability of silicon-fabrication facilities that are used in micro- and nanoelectronics.

Thanks to intense research activity over the past 15 years, there have been many breakthroughs in the field of silicon photonics. Examples include hybrid silicon lasers³, various types of modulator⁴ (devices that convert electronic information into optical signals), high-speed light detectors⁵ (photodetectors) and complex optoelectronic circuits⁶. Several companies currently sell products based on silicon photonics chips, and many more are poised to do so in the near future.

In the electronics industry, complementary-metal-oxide-semiconductor (CMOS) technology is used to create computer processors and memory, communication chips and image sensors. This technology is based on silicon and depends on the ability to cram a large number of transistors and electronic circuits

on to a single chip. Similarly, the integration of large numbers of electronic and photonic circuits on a single chip is crucial for meeting the requirements of computer processors and communication links in data centres, in terms of data-transmission rates, power consumption, scalability and complexity.

The main challenge for such integration has been the incompatibility of the material platforms used in silicon electronics and photonics. CMOS technology uses either bulk silicon substrates or thin silicon-on-insulator wafers⁶. The former is by far the most dominant platform because of its abundant supply chain and low cost. By contrast, silicon photonics usually requires thick silicon-on-insulator wafers that have a limited supply chain and are too expensive for many applications, such as computer memory. A long-term goal has therefore been to integrate electronic and photonic components using standard CMOS-manufacturing techniques and material platforms, without affecting the performance of such components.

Atabaki and colleagues have made a breakthrough in this regard by decoupling the formation of photonic devices from that of transistors, and by successfully incorporating these photonic devices into bulk silicon CMOS chips. The authors used standard CMOS-manufacturing methods, and introduced only a few changes to the fabrication process to create areas dedicated to photonic devices in the bulk silicon. The devices were integrated during the processing of the transistors. This involved the addition of isolated patches (islands) of the insulator material silicon dioxide to the bulk silicon and the deposition of a thin film of polycrystalline silicon on

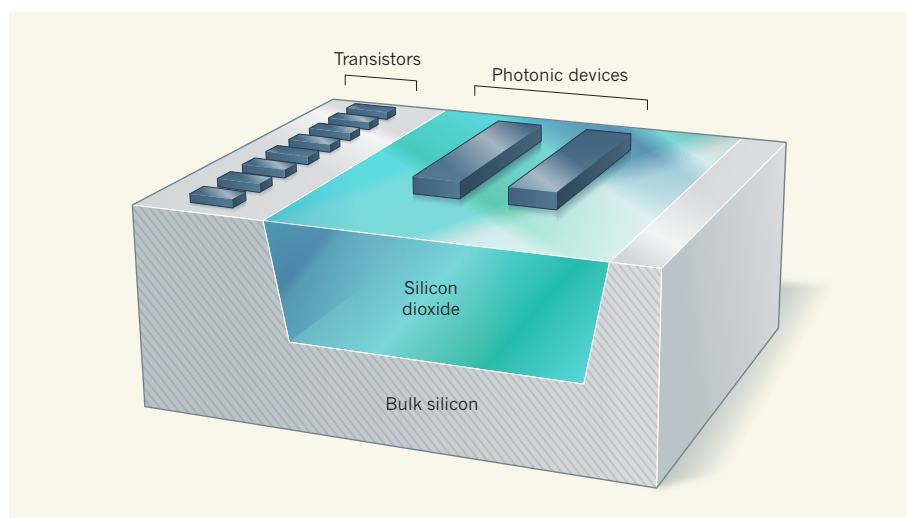


Figure 1 | Optoelectronic integration. Atabaki *et al.*¹ report a technique for integrating electronic and photonic devices on a single silicon microchip. The authors added isolated patches (islands) of the insulator material silicon dioxide to a bulk silicon substrate — for simplicity, a single island is shown here. They then deposited a thin film of polycrystalline silicon on top. Photonic devices and electronic devices known as transistors were fabricated from this film; the former in the silicon-on-insulator region and the latter in the bulk silicon. (Adapted from Fig. 1b of ref. 1.)

top. Photonic devices were fabricated in this silicon-on-insulator region, whereas transistors were formed in standard bulk silicon regions on the CMOS chip (Fig. 1).

Although the electronic and photonic properties of crystalline silicon are superior to those of polycrystalline silicon, because the former has a more uniform structure, it is not possible to grow crystalline silicon on top of silicon dioxide. Atabaki *et al.* therefore opted for polycrystalline silicon, which is relatively cheap and readily available because it is used in transistor fabrication. The authors used this material to create various photonic components, including waveguides (structures that enable light propagation on chips), optical filters known as micro-ring resonators, vertical grating couplers (for coupling light between waveguides and optical fibres), high-speed modulators and photodetectors.

The performance of these components was similar to or better than previous demonstrations in polycrystalline silicon^{7–9}. But, more importantly, the performance was unaffected by the fact that such components operated next to electronic-circuit blocks composed of millions of CMOS transistors. Consequently, the authors' silicon-photonics chips can achieve many of the goals of systems that require multiple chips, with substantial cost, scalability and performance advantages.

Atabaki and colleagues' results are impressive, but there are several aspects that could be improved. For example, optical loss in the waveguides could be reduced, and the filtering of light in the micro-ring resonators and the coupling efficiency of the vertical grating couplers could be increased. The authors suggest that optical loss might be minimized by refining the polishing process that they used

to reduce the roughness of the silicon dioxide islands and the polycrystalline-silicon film. Such an improvement would lead to chips that have better photodetector sensitivity, lower voltage requirements and lower power consumption — all of which are crucial for the realization of efficient on-chip optoelectronic systems.

Further optimization of the polycrystalline-silicon film could enhance the speed of the modulators and photodetectors, which is paramount for future optical connections that can transmit data at rates of multi-terabytes per second. Atabaki *et al.* fabricated their silicon photonics chips using a technology based on 65-nanometre transistors, and it will be interesting to see whether their approach can be extended to smaller scales at which an even greater density of transistors can be integrated. Future work could also examine how the approach could be used for optical connections inside microprocessors.

Although there are several challenges to be overcome, the authors' work is a milestone on the path towards the mass production of on-chip optoelectronic systems. We can expect an exciting period of development of such systems, and their demonstration for a host of applications. In the future, they might be as ubiquitous as today's electronic microchips. ■

Goran Z. Mashanovich is in the Optoelectronics Research Centre, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton SO17 1BJ, UK. e-mail: g.mashanovich@soton.ac.uk

1. Atabaki, A. H. *et al.* *Nature* **556**, 349–354 (2018).
2. Rickman, A. *Nature Photon.* **8**, 579–582 (2014).
3. Zhou, Z., Yin, B. & Michel, J. *Light Sci. Appl.* **4**, e358 (2015).



50 Years Ago

A modern approach to the study of salmon migration has been started ... The project involves the use of a computer in a three year analysis of the factors involved in the migration of salmon to fresh-water rivers and those which might affect the fish on its return to the sea. The analysis appears to be the first of its kind in Europe and possibly in the world ... Data from a number of rivers will be fed into the computer at intervals of about a month. Physical factors which may be connected with the movement of salmon are being recorded, particularly river flow, air and water temperature, amount of light ... and solar radiation ... In addition, special fish traps record all the fish swimming up or down the river.
From Nature 20 April 1968

100 Years Ago

The possibility of an aerial mail has often been commented upon ... and it is very interesting to note that a company has actually been formed in Norway for the purpose of establishing a mail service between Aberdeen and Stavanger. This trip was made just before war broke out ... in about five hours' flying, and it is estimated that the mail services will reduce this to four and a half hours with modern machines. An extension of the system to Christiania and Copenhagen is contemplated, and it is hoped that letters leaving Aberdeen in the morning would be delivered in both these cities in the afternoon ... The value of such a mail service would be very great at a time when the overseas service is so seriously hampered by the German submarine campaign, and the satisfactory establishment of the contemplated Norwegian service would undoubtedly soon lead to a general use of the aeroplane for rapid international communication.
From Nature 18 April 1918

4. Thomson, D. *et al.* *J. Opt.* **18**, 073003 (2016).
5. Vivien, L. *et al.* *Opt. Express* **20**, 1096–1101 (2012).
6. Sun, C. *et al.* *Nature* **528**, 534–538 (2015).
7. Sun, C. *et al.* *IEEE J. Solid-State Circuits* **50**, 828–844 (2015).
8. Preston, K., Manipatruni, S., Gondarenko, A., Poitras, C. B. & Lipson, M. *Opt. Express* **17**, 5118–5124 (2009).
9. Mehta, K. K. *et al.* *Opt. Lett.* **39**, 1061–1064 (2014).

MICROBIOLOGY

A stockpile of antiviral defences

The full list of weapons used by bacteria against viruses is not known. A computational approach has uncovered nine previously unidentified antiviral systems, encoded by genes near known defence genes in bacterial genomes.

SÉBASTIEN LEVESQUE & SYLVAIN MOINEAU

Fighting viruses is no easy task. Bacteria have survived attacks by viruses called phages by evolving sophisticated defence strategies that enable them to thrive even in virus-rich ecosystems. However, phages have evolved counter-tactics to thwart such mechanisms¹, leading to a biological arms race. Now Doron *et al.*² report the identification of previously unknown antiviral systems in bacteria.

Anti-phage systems usually target key steps in viral replication. For example, some systems prevent phage binding to bacterial cells, whereas others block entry of the viral genome

into the cell³. Certain bacterial proteins can halt intracellular phage replication^{3,4}. Although this often leads to the death of infected cells, it can protect neighbouring cells from infection. Perhaps the best known anti-phage systems are restriction enzymes and CRISPR–Cas. These two systems^{5–7}, can cleave non-host DNA in a sequence-specific manner, and have also been widely adapted as molecular tools in the biological sciences.

As knowledge of the diversity of Earth's viruses has grown⁸, along with the potential of using such information to develop further biotechnology tools, investigation into anti-phage systems has surged. Many lines of evidence have indicated that the list

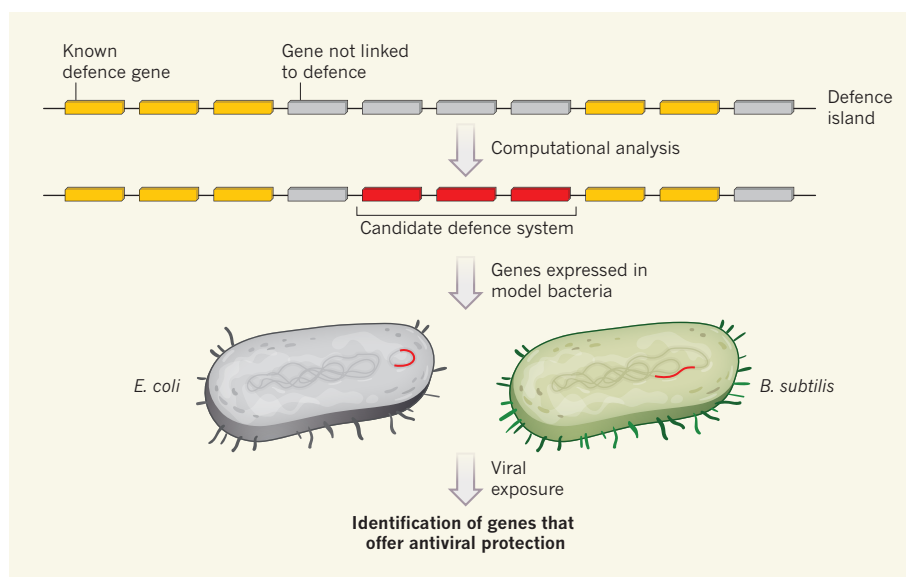


Figure 1 | Identifying antiviral systems in bacteria. Bacterial defence genes (yellow) are found in regions of the genome known as defence islands. Doron *et al.*² sought to identify more such genes by analysing genes within these islands that had not previously been linked to defence functions (grey). They used computational analysis involving a range of criteria, including whether the genes were located in defence islands in many different types of bacterium. The authors also identified neighbouring genes that might function together as a defence system. These proposed defence-system genes were then expressed in the model bacteria *Escherichia coli* and *Bacillus subtilis*. The bacteria were exposed to various viruses to test whether the genes offered protection against infection. The authors confirmed that nine of the defence systems they tested had antiviral functions.

of microbial-defence ‘weapons’ is probably far from complete.

Previous computational analyses have shown that defence genes cluster together in bacterial genomes in specific regions called defence islands⁹. Enter Doron *et al.*, armed with the knowledge⁹ that these regions also contain many gene families that have unknown functions. The authors analysed more than 45,000 microbial genomes to find genes that are frequently found in defence islands. For their analysis, they grouped the encoded proteins into families that share a specific structural domain. Doron and colleagues analysed 14,083 protein families, and focused on those in which at least 65% of the encoding genes were located near known defence systems. These genes were then used as ‘anchors’ from which to investigate neighbouring genes, because defence genes are often found to be part of a series of consecutive genes that function together in the same defence process.

The authors pinpointed 335 families of interest. After further studies to identify gene clusters that are evolutionarily conserved across multiple genomes and in a broad distribution of microbes, they selected 28 such clusters for functional testing. They expressed the genes in two model bacteria: *Bacillus subtilis* and *Escherichia coli* (Fig. 1). In *B. subtilis*, the selected genes were integrated into the genome, whereas in *E. coli*, they were engineered into circular-plasmid DNA.

The bacteria successfully expressed at least one example of 26 of these candidate defence systems, as confirmed by RNA sequencing. They also expressed six known defence systems as controls. The bacteria were then exposed to a range of phages belonging to four distinct phage families known to infect them. Remarkably, nine of the 26 systems offered protection against at least one phage. These defence systems contained up to five genes. One system was present in 3% of the bacterial genomes analysed, and another was found in 4% of microbes investigated. The authors named the systems after mythological protective deities.

Some selected candidates had no anti-phage activity. This was not surprising, because they were tested under specific laboratory conditions and were expressed in hosts that do not normally express these genes: defence mechanisms are often effective only against specific phage groups. Indeed, only three of the six known defence systems used as controls provided protection against phages in the experiments. The authors speculated that some of the defence systems they had identified might specifically defend against plasmid introduction. In an experiment testing the efficiency of plasmid introduction into *B. subtilis*, they found that the presence of one of the defence systems substantially reduced the level of plasmid introduction. Altogether, the authors identified ten defence systems (nine antiviral and one antiplasmid) in various microbes.

Doron and colleagues proposed distinct modes of action for some of these defence mechanisms on the basis of the presence of specific domains in some of the bacterial proteins. For example, one protein has a TIR domain. This domain is a key component of the innate immune system of mammals, plants and invertebrates and it functions in signalling pathways activated in response to the recognition of infectious agents. However, in-depth mechanistic studies are needed to draw any

“Defence genes cluster together in bacterial genomes in specific regions called defence islands.”

conclusions about how these newly identified defence systems might function. The discovery of this hidden stockpile of anti-phage weapons is exciting, and emphasizes the fact that the complete array of bacterial defence systems remains unknown. Doron and colleagues’ experiments might even have missed some systems because of the technical methods they used. For example, some groups of genes tested might have been incompatible with the model bacteria used, or might provide protection only against phages that weren’t tested. Indeed, the recent discovery of a major lineage of marine viruses¹⁰ is a reminder that our inventory of viruses continues to expand.

The authors have convincingly demonstrated an effective computational approach for discovering bacterial defence systems. The

presence of multiple such mechanisms in a given bacterium gives the microbe a robust safeguard against viral infection¹¹, so the decision to investigate defence islands was an astute one. In the never-ending battle between phages and bacteria, it will also be interesting to learn how phages have evolved to neutralize or circumvent these newly unmasked weapons. Rest assured, phages are here to stay, and are bound to mount a counter-attack. ■

Sébastien Levesque and Sylvain Moineau are in the Department of Biochemistry, Microbiology and Bioinformatics, Faculty of Sciences and Engineering, Laval University, Quebec City, Quebec G1V 0A6, Canada. S.M. is also at the Félix d'Hérelle Reference Center for Bacterial Viruses, Laval University. e-mail: sylvain.moineau@bcm.ulaval.ca

1. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. *Nature Rev. Microbiol.* **11**, 675–687 (2013).
2. Doron, S. *et al. Science* **359**, eaar4120 (2018).
3. Labrie, S. J., Samson, J. E. & Moineau, S. *Nature Rev. Microbiol.* **8**, 317–327 (2010).
4. Snyder, L. *Mol. Microbiol.* **15**, 415–420 (1995).
5. Pingoud, A., Wilson, G. G. & Wende, W. *Nucleic Acids Res.* **42**, 7489–7527 (2014).
6. Barrangou, R. *et al. Science* **315**, 1709–1712 (2007).
7. Jinek, M. *et al. Science* **337**, 816–822 (2012).
8. Paez-Espino, D. *et al. Nature* **536**, 425–430 (2016).
9. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. *J. Bacteriol.* **193**, 6039–6056 (2011).
10. Kauffman, K. M. *et al. Nature* **554**, 118–122 (2018).
11. Dupuis, M.-É., Villion, M., Magadán, A. H. & Moineau, S. *Nature Commun.* **4**, 2087 (2013).

This article was published online on 16 April 2018.

ASTROPHYSICS

Bounteous black holes at the Galactic Centre

X-ray observations have revealed a dozen stellar-mass black holes at the centre of the Galaxy, implying that there are thousands more to be found. The discovery confirms a fundamental prediction of stellar dynamics.

MARK R. MORRIS

A dense cluster of stars surrounds the supermassive black hole that lies at the Galactic Centre. Stars that live and die in the cluster are almost always held captive by the irresistible gravity of this strong concentration of mass. Consequently, the black-hole remnants left behind by the deaths of massive stars are predicted to have piled up in the central parsec (3.26 light years) of the Galaxy during its lifetime. Theoretical estimates of the number of stellar-mass black holes in this region range from the thousands to the tens of thousands^{1–3}. Writing in a previous issue of *Nature*, Hailey *et al.*⁴ reported on

what could be the first observational evidence for such a black-hole cluster.

All stars emit X-rays, but only the brightest stellar X-ray sources at the centre of the Galaxy can be observed. Nevertheless, with a single field of view pointing towards the Galactic Centre, the Advanced CCD Imaging Spectrometer (ACIS) of NASA’s space-based Chandra X-Ray Observatory has detected thousands of these sources. Almost all are found in close binary systems that comprise a normal star and a compact companion. The X-rays are generated by gas that is subjected to strong heating when it is pulled out of the normal star and transferred (accreted) onto, or into, its companion.

Most of the X-ray sources are binaries that contain a white dwarf as the companion. Such systems are known as cataclysmic variables because their accretion flows lead to an accumulation of matter on the surface of the white dwarf that then undergoes violent episodes of nuclear burning. Much less common at the Galactic Centre are binaries in which the companion is a neutron star or a black hole. These systems are referred to as low-mass X-ray binaries (LMXBs) because of the relatively low mass of the normal star that they contain. High-mass X-ray binaries, in which the normal star is massive, highly luminous and can be seen easily using infrared surveys, have been ruled out through observation^{5,6} in the central region of the Galaxy considered by Hailey and colleagues.

A compensating factor for the usual scarcity of black-hole LMXBs at the Galactic Centre is the phenomenon of mass segregation. In this process, the gravitational interactions of the stars that orbit the Galactic Centre cause the heaviest ones, or binary stars, to move closer to the centre and the lightest ones to migrate outwards^{1–3}. Stellar-mass black holes typically have masses that are 5 to 15 times that of the Sun⁷ — much greater than those of most other stars in this environment. Such black holes should therefore become strongly concentrated at the Galactic Centre, regardless of whether they are isolated or part of binary systems. Neutron stars, which usually have masses of 1 to 2 solar masses^{7,8}, should be much less concentrated.

Hailey *et al.* used a broad-brush examination of the spectra of X-ray sources in the Galactic Centre to distinguish between LMXBs and the more abundant cataclysmic variables. The latter have spectra that are characteristic of thermal emission processes, including prominent spectral lines associated with iron, whereas the former have non-thermal, featureless spectra, indicating emission from extremely high-velocity particles.

The authors distinguished between neutron-star and black-hole LMXBs by using the fact that neutron-star LMXBs undergo violent outbursts of X-rays on timescales shorter than the 18 years for which Chandra has been monitoring the Galactic Centre. By contrast, outbursts from black-hole LMXBs recur on much longer timescales, so the chance that a particular one will have undergone an outburst during the observation window is small. More than a dozen outbursting

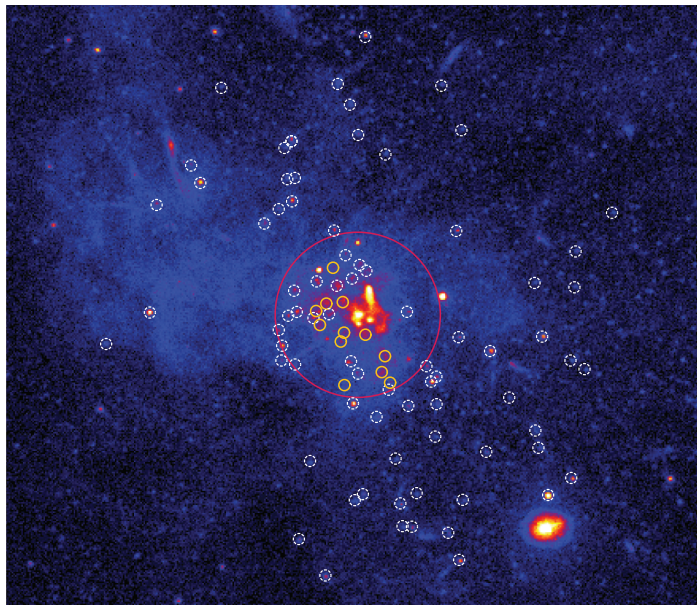


Figure 1 | X-ray emission from the Galactic Centre. Hailey *et al.*⁴ examined the spectra of the brightest stellar X-ray sources at the centre of the Galaxy. The locations of these sources are indicated by small circles. The authors identified 12 sources (yellow circles) that have the expected characteristics of close binary systems comprising a low-mass star and a black hole. Such sources are contained in the Galaxy's central parsec (3.26 light years), which is indicated by the red circle. At the core of this region lies a supermassive black hole, which is itself a prominent X-ray source. The background colours represent the strength of the X-ray emissions, from low (black) to high (yellow).

neutron-star LMXBs (also known as X-ray transients) have been discovered in the Galactic Centre, spread out far beyond the central parsec, and this might be almost the entire population of such objects⁹.

After Hailey and colleagues had accounted for the cataclysmic variables and neutron-star LMXBs, there remained 12 X-ray sources with the expected characteristics of black-hole LMXBs, all of which were located in the central parsec (Fig. 1). This result provides strong evidence to support the hypothesis that black holes are concentrated at the Galactic Centre. Of course, this includes only the close binary systems; there is probably a much larger population — perhaps as many as 10,000 — of isolated, and presently unobservable, black holes in the same volume. But such extrapolation is difficult because the effectiveness of the various mechanisms for producing close binaries is uncertain (but see ref. 10).

The lifetimes of close binaries in such an environment are also uncertain. For instance, two known effects can cause the members of such a system to eventually coalesce into a single object. In the first, close gravitational encounters with other stars cause the distance between the members of the binary to decrease until the pair merges. And in the second, on a shorter timescale, the supermassive black hole at the centre of the Galaxy, around which all binary systems in the region orbit, drives mergers. This occurs because the gravity of the supermassive black hole gradually increases the eccentricity of the orbits of the

stars in the binary. Eventually, these orbits become so elongated that the two members make contact and undergo a relatively violent coalescence^{11,12}.

A merging black-hole LMXB would result in a black hole of increased mass. If this new object formed another binary system that then also merged, and such a chain of events continued, it would be possible to produce black holes with masses of up to several tens of times that of the Sun¹³. Such masses lie in the range that has been determined to account for the detailed gravitational-wave signatures of merging binaries that contain black holes¹⁴. It is unclear whether such large-mass black holes can be created in single supernova explosions of extremely massive stars, but Hailey and colleagues' findings pave the way towards understanding not only how such black holes can be created, but also how they end up in binary systems.

The next set of observations will probably be a long time coming because Hailey *et al.* have already used much of Chandra's existing database for their analysis. In the

near future, theoretical investigations of the dynamical formation and evolution of binary systems will be crucial for understanding central clusters of black holes that could be common in galaxies. ■

Mark R. Morris is in the Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, California 90095-1547, USA.
e-mail: morris@astro.ucla.edu

1. Miralda-Escudé, J. & Gould, A. *Astrophys. J.* **545**, 847–853 (2000).
2. Morris, M. *Astrophys. J.* **408**, 496–506 (1993).
3. Freitag, M., Amaro-Seoane, P. & Kalogera, V. *Astrophys. J.* **649**, 91–117 (2006).
4. Hailey, C. J. *et al. Nature* **556**, 70–73 (2018).
5. Mauerhan, J. C. *et al. Astrophys. J.* **703**, 30–41 (2009).
6. Laycock, S. *et al. Astrophys. J.* **634**, L53–L56 (2005).
7. Raithe, C. A., Sukhbold, T. & Özel, F. *Astrophys. J.* **856**, 35 (2018).
8. Özel, F. & Freire, P. *Annu. Rev. Astron. Astrophys.* **54**, 401–440 (2016).
9. Degenaar, N. *et al. Astron. Astrophys.* **545**, A49 (2012).
10. Genovese, A., Stone, N. C., Metzger, B. D. & Ostriker, J. P. Preprint at <https://arxiv.org/abs/1804.01543> (2018).
11. Naoz, S. *Annu. Rev. Astron. Astrophys.* **54**, 441–489 (2016).
12. Stephan, A. P. *et al. Mon. Not. R. Astron. Soc.* **460**, 3494–3504 (2016).
13. Antonini, F. & Rasio, F. A. *Astrophys. J.* **831**, 187 (2016).
14. Abbott, B. P. *et al. Phys. Rev. Lett.* **118**, 121101 (2017).

This article was published online on 16 April 2018.

Laser spectroscopic characterization of the nuclear–clock isomer $^{229\text{m}}\text{Th}$

Johannes Thielking¹, Maxim V. Okhapkin¹, Przemysław Glowacki^{1,6}, David M. Meier¹, Lars von der Wense², Benedict Seiferle², Christoph E. Düllmann^{3,4,5}, Peter G. Thirolf² & Ekkehard Peik^{1*}

The isotope ^{229}Th is the only nucleus known to possess an excited state $^{229\text{m}}\text{Th}$ in the energy range of a few electronvolts—a transition energy typical for electrons in the valence shell of atoms, but about four orders of magnitude lower than typical nuclear excitation energies. Of the many applications that have been proposed for this nuclear system, which is accessible by optical methods, the most promising is a highly precise nuclear clock that outperforms existing atomic timekeepers. Here we present the laser spectroscopic investigation of the hyperfine structure of the doubly charged $^{229\text{m}}\text{Th}$ ion and the determination of the fundamental nuclear properties of the isomer, namely, its magnetic dipole and electric quadrupole moments, as well as its nuclear charge radius. Following the recent direct detection of this long-sought isomer, we provide detailed insight into its nuclear structure and present a method for its non-destructive optical detection.

^{229}Th has a low-energy transition between the nuclear ground state and a long-lived isomer, $^{229\text{m}}\text{Th}$, at an excitation energy of about 7.8 eV. This enables the application of precision laser spectroscopy methods to excite and detect this nuclear transition^{1–3}. A nuclear clock—that is, an optical clock that uses this low-energy transition as the frequency reference—is expected to benefit from the smaller sensitivity of the nucleus to external perturbations, including frequency shifts from electromagnetic fields, compared to electronic transitions exploited in current atomic clocks^{4,5}. To achieve this application, we studied the nuclear properties of ^{229}Th , as well as the excitation of the isomeric state and methods for its non-destructive observation⁶.

The nuclear structure of ^{229}Th has been studied via γ -ray spectroscopy of the radiation emitted after the α decay^{7–9} of ^{233}U , after the β decay^{10,11} of ^{229}Ac , after Coulomb excitation¹², and through a (d,t) transfer reaction¹³ with ^{230}Th . The low-lying levels can be assigned to rotational bands described by the Nilsson model¹⁴. The ^{229}Th ground state is the bandhead of a $5/2^+$ [633] rotational band and its nuclear moments have been determined experimentally^{12,15}. A second rotational band has been identified as $3/2^+$ [631]; its bandhead—the low-energy isomer $^{229\text{m}}\text{Th}$ that is investigated here—is still unresolved by γ spectroscopy.

The transition energy between the ground state and the isomer has been determined^{16,17} indirectly, as the difference between the γ -ray energies of intraband and interband transitions, to be $E_{\text{is}} = 7.8(5)$ eV (all uncertainties represent a 68% confidence level). This corresponds to ultraviolet radiation of wavelength 160(10) nm, where the uncertainty is about 17 orders of magnitude larger than the expected natural linewidth. Depending on the electronic structure that surrounds the nucleus, the isomer may decay quickly via internal conversion^{2,3} or radiatively with an estimated natural (that is, unperturbed) lifetime of a few thousands of seconds^{11,18–20}. The nuclear moments of the isomer have been estimated from nuclear structure models^{18,21,22}. Many experimental attempts to induce and detect an optical excitation of this isomer have failed, impeded by the difficulty of producing widely tunable intense vacuum-ultraviolet radiation, by the background of ionizing radiation from the ^{229}Th samples and by competition with non-radiative relaxation processes^{6,23,24}. Apart from the spectroscopic

determination of the nuclear spin and indirect measurements of the excitation energy^{7,8,16,17}, no experimental data on the nuclear properties of the isomer have been available until recently. Using recoil ions from the decay of ^{233}U as a source of $^{229\text{m}}\text{Th}$, electrons emitted from the internal-conversion decay of the isomer in neutral thorium were detected²⁵ and the half-life for this process was measured²⁶.

The availability of the isomer through recoil ions provides a way to measure the unknown nuclear properties of $^{229\text{m}}\text{Th}$ via laser spectroscopy of electronic transitions. Here we report the optical detection of ions in the $^{229\text{m}}\text{Th}$ isomeric state and of the resolved hyperfine structure (HFS), which arises from the interaction of the isomer nucleus with the valence electrons (see Methods).

Laser spectroscopy of trapped Th^{2+} ions

Of the charge states Th^+ , Th^{2+} and Th^{3+} that are extracted from a ^{233}U source, Th^{2+} was selected for the experiments because of its high production yield from the recoil source, the long lifetime of the isomer²⁵ and its convenient electronic-level structure, which enables hyperfine spectroscopy with diode lasers with background-free fluorescence detection in the visible spectral range.

For high-resolution spectroscopy of the HFS of $^{229}\text{Th}^{2+}$, we use two independent linear radio-frequency ion traps^{25,27} (see Fig. 1a and b and Methods). One of the traps (located at the Physikalisch-Technische Bundesanstalt, PTB) is loaded with Th^+ produced by laser ablation from a target containing ^{229}Th and ^{232}Th . Three-photon ionization of trapped Th^+ is used to produce Th^{2+} . The second trap (located at Ludwig-Maximilians-Universität, LMU) is loaded with ^{229}Th recoil ions from the α decay of ^{233}U , where the isomeric state is populated via a 2% decay branch⁹ (see Methods for details about the generation of the $^{229\text{m}}\text{Th}^{2+}$ ion beam). Therefore, the trapped ion cloud consists of a mixture of ions in the ground and the isomeric nuclear states. Daughter products of the ^{233}U decay chain are also trapped, but do not disturb the spectroscopic measurement (see Methods for details). The combination of the measurements in both traps allows us to identify clearly the hyperfine components of $^{229\text{m}}\text{Th}$, which appear only with the trapped recoil ions, and to measure the isotope and isomer shifts. In both traps, the ions are cooled to near room temperature

¹Physikalisch-Technische Bundesanstalt, Braunschweig, Germany. ²Ludwig-Maximilians-Universität München, Garching, Germany. ³GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany. ⁴Helmholtz-Institut Mainz, Mainz, Germany. ⁵Johannes Gutenberg-Universität, Mainz, Germany. ⁶Present address: Poznań University of Technology, Poznań, Poland.

*e-mail: ekkehard.peik@ptb.de

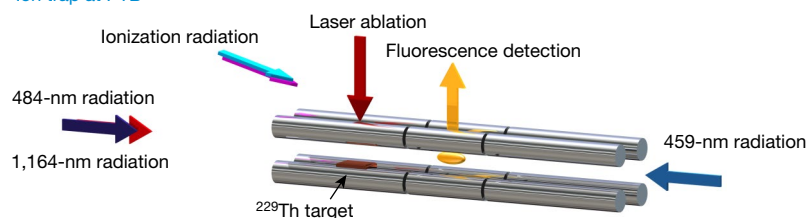
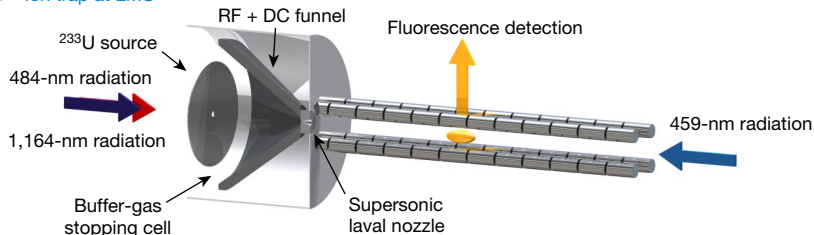
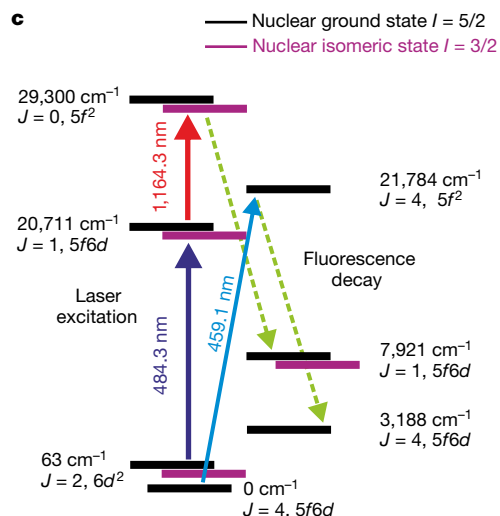
a Ion trap at PTB**b** Ion trap at LMU

Fig. 1 | Experimental setup and $^{229}\text{Th}^{2+}$ level scheme. **a**, Schematic of the configuration of the ion source, ion trap and laser beams at PTB. **b**, Corresponding configuration at LMU. RF, radio frequency; DC, direct current. **c**, Transitions and electronic configurations of Th^{2+} levels relevant

c

to the experiment, labelled by their energy in cm^{-1} and the electronic angular momentum J , which are the same for both nuclear states (see Methods for details). Solid arrows indicate laser excitation and dashed lines show fluorescence decay channels.

by collisions with a high-purity buffer gas (argon at PTB, helium at LMU; see Methods). Recording an excitation spectrum by scanning a laser across the Doppler-broadened lines yields a resolution of about 700 MHz. This does not allow us to resolve the HFS of $^{229}\text{Th}^{2+}$ lines and to distinguish the resonances of $^{229}\text{Th}^{2+}$ from those of $^{232}\text{Th}^{2+}$. As an example, Fig. 2b shows a line of ^{232}Th (nuclear spin $I = 0$, no HFS splitting) and an unresolved HFS lineshape of ^{229}Th . For higher resolution we use two-step laser excitation, which is free from Doppler broadening^{28,29}. In this approach, the first laser excites ions of a narrow velocity class out of the thermal distribution to an intermediate state, where they are probed by resonant excitation to a higher-lying level using a second tunable laser. These ions are detected by a sensitive fluorescence detection method using decay channels at other wavelengths, which are free from stray background laser light.

For the two-step laser excitation we choose the transition from the 63_2 electronic state to 29300_0 via the 20711_1 intermediate state (the states are labelled by their energy in cm^{-1} and the electronic angular momentum J as subscript, as shown in Fig. 1). The population of the 63_2 state is in equilibrium with the electronic ground state (0_4), at a ratio of approximately 0.4, through collisions with the buffer gas. The analysis of the isomeric HFS is simple because the two-step excitation with electronic angular momentum $2 \rightarrow 1 \rightarrow 0$ (Fig. 1) leads to a small number of HFS components. Nine hyperfine resonances are present for the nuclear ground state ($I = 5/2$) and eight hyperfine resonances are expected to appear⁷ for the isomer with nuclear spin $I = 3/2$ (see Extended Data Fig. 1).

The spectroscopy part of the experimental setup consists of three external-cavity diode lasers (ECDLs). The two-step excitation is provided by ECDLs at wavelengths of 484 nm and 1,164 nm, with overlapping beams aligned along the trap axis. The third diode laser, at 459 nm, is used for single-photon excitation of Th^{2+} from the 0_4 ground state to the 21784_4 state to monitor the number of Th^{2+} ions in the trap and thus normalize the fluorescence signals observed from the different HFS components. This is required because the ion number decreases with time owing to chemical reactions and charge exchange with impurities in the buffer gas. The laser setup is depicted in Extended Data Fig. 2 and details are given in Methods.

Figure 2b shows the single-photon excitation spectrum of the transition from 63_2 to 20711_1 , obtained by scanning the frequency of the 484-nm ECDL over the resonances of ^{232}Th and ^{229}Th (see Methods). These data allow us to measure the isotopic shift and to determine the search range of the HFS of the isomer. The spectrum in Fig. 2a

shows individual HFS components of $^{229}\text{Th}^{2+}$, obtained in the second excitation step by scanning the 1,164-nm ECDL over the $20711_1 \rightarrow 29300_0$ transition. The two-step resonances are mapped for a systematic search for the unknown frequencies of the isomer resonances and for the quantitative analysis of the HFS spectra. The frequency of the 484-nm laser is tuned within the Doppler-broadened HFS of the $63_2 \rightarrow 20711_1$ line in 35 discrete steps of about 120 MHz. At each frequency step, the 1,164-nm laser is scanned continuously over a frequency range of more than 4 GHz to detect the two-step HFS resonances. The frequency of the first-step laser is stabilized at each position of the map using a Fizeau wavemeter, resulting in absolute instability of ≤ 5 MHz. The full-width at half-maximum of the two-step resonances obtained in the PTB trap (which uses argon as a buffer gas) is 70 MHz and 40 MHz for those observed at LMU (helium buffer gas). The two-step excitation mapping is performed twice, with co- and counter-propagating laser beams, to confirm the identification of the HFS components.

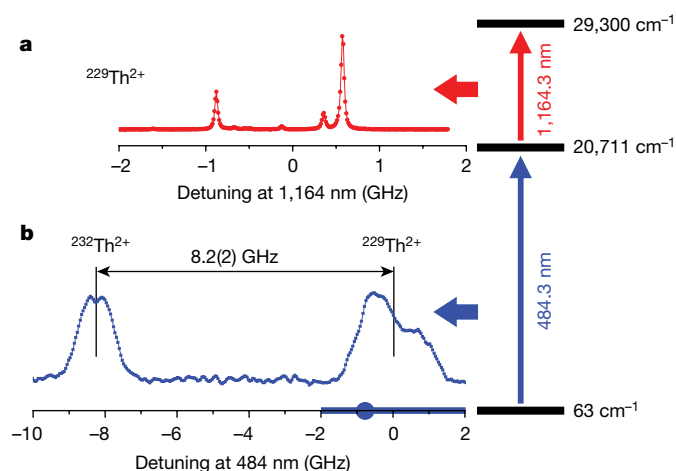


Fig. 2 | Two-step excitation spectra. **a**, Example of a high-resolution two-step spectrum. **b**, Doppler-broadened single-photon spectrum of the first step. The single-step excitation shows an isotopic shift of 8.2(2) GHz between $^{232}\text{Th}^{2+}$ and $^{229}\text{Th}^{2+}$. For the two-step spectrum, the frequency of the first laser is fixed at -800 MHz detuning (indicated by the blue dot) with respect to the ^{229}Th HFS centre. Measurements are performed at 35 discrete steps of about 120 MHz with the 484-nm laser, within the range indicated by the blue bar on the frequency axis.

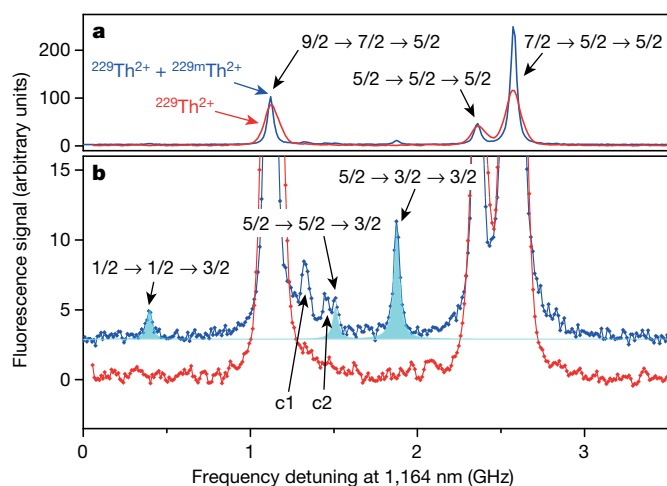


Fig. 3 | Comparison of excitation spectra measured in the PTB and LMU traps. **a**, Two-step excitation resonances of Th^{2+} , obtained with the first laser stabilized at -800 MHz detuning with respect to the ^{229}Th HFS centre and with the second laser being scanned. Red and blue colours indicate data acquired in the PTB and the LMU trap, respectively. 2% of the ions in the LMU trap are in the isomeric state. **b**, Magnified view of **a**, with the LMU signal up-shifted for easier inspection. The arrows indicate the total momenta F and F^m of the transitions of the ground-state (**a**) and isomeric (**b**) resonances (Extended Data Fig. 1), respectively, and the isomeric-state peaks are shaded in cyan. Resonances c1 and c2 belong to the nuclear ground state and arise from collision-induced changes of the intermediate-state HFS (Extended Data Table 1); the transitions involved have total momenta F of $7/2 \rightarrow (5/2, 7/2) \rightarrow 5/2$ and $7/2 \rightarrow (5/2, 3/2) \rightarrow 5/2$, respectively, where the two F numbers in parentheses are mixed by collisions.

All spectra are available in the Supplementary Information (Supplementary Figs. 1–29 and 31–59). Because the expected isomer signal is only about 2% of the signal from ions in the ground state, we choose averaging times of typically four hours per spectrum.

Detection of the isomeric HFS

In the PTB trap we detect all nine resonances of the HFS of the ground state. A typical spectrum of the nuclear ground-state HFS resonances for co-propagating beams, obtained at PTB with two-step laser excitation, is shown in Fig. 3 with red points. The second signal (blue points) shows the HFS signal obtained in the LMU trap, where a small fraction of the ions is in the isomeric state. The HFS resonances of the isomeric state are clearly observed when compared with the data acquired at PTB. Figure 4 shows four resonances of the HFS of the isomeric state in a logarithmic scale for a different frequency of the first-step 484-nm laser.

In total, we observe seven out of eight resonances of $^{229m}\text{Th}^{2+}$ in both (co- and counter-propagating) beam configurations (see Extended Data Fig. 3). The amplitude of the eighth resonance is calculated to be small with respect to the signal-to-noise ratio achieved in the experiment. The fraction of the ions in the isomeric state, obtained from the ratio of integrated fluorescence signals of the isomeric and ground-state resonances, is determined to be 2.1(5)% (see Methods and Extended Data Fig. 4). This confirms the previously assumed branching to the isomeric state, which was inferred from γ spectroscopy⁹.

We observe a collision-induced change of the intermediate-state HFS population due to interactions with the buffer gases, such as the resonances c1 and c2 in Fig. 3. This effect is more substantial for He in the LMU trap than for Ar in the PTB trap. The positions of these resonances can be calculated from the measured hyperfine splitting of the intermediate state, enabling us to identify them in the spectra. This prevents the wrong assignment of resonances originating from collisions as isomeric-state HFS peaks. The width of these resonances is approximately 1.5 times larger than those of the nine main HFS resonances.

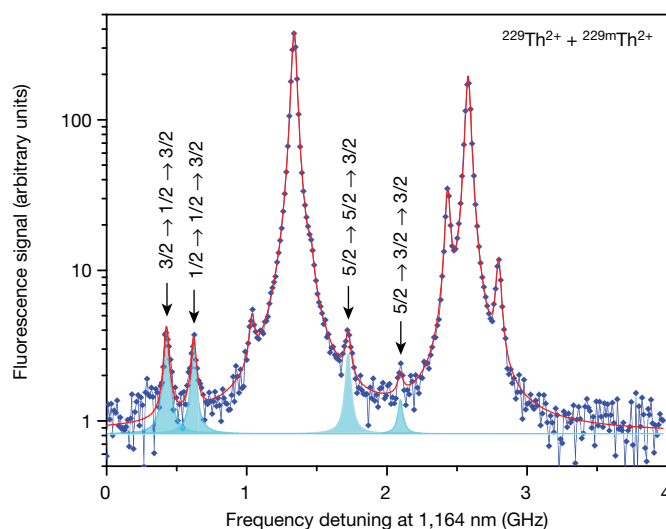


Fig. 4 | HFS resonances of nuclear isomeric and ground states. Two-step excitation resonances of the nuclear isomer HFS are displayed (shaded in cyan), showing the relative strengths (in logarithmic scale) and frequency range of the isomeric- and ground-state resonances (Extended Data Fig. 1). The first laser is stabilized at -260 MHz detuning with respect to the ^{229}Th HFS centre and the second laser is scanned. The unlabelled peaks correspond to the ground state.

Furthermore, the amplitude ratio between these resonances and the direct two-step resonances drops considerably with the reduction of the He buffer-gas pressure, indicating that these resonances are indeed caused by collisions (see Extended Data Fig. 5).

Isomer properties

The observation of the isomeric-state HFS allows us to determine the magnetic dipole and quadrupole moments and the nuclear charge radius. To derive these properties, we determine the hyperfine constants A (magnetic dipole) and B (electric quadrupole) of the electronic structure for both the ground and isomeric nuclear states. We measure the frequency intervals between the resonances relative to the transmission peaks of a reference cavity and use a least-squares fit to determine the hyperfine constants for the 63_2 and the 20711_1 electronic states (see Methods). The upper electronic state of the two-step excitation has $J = 0$ and therefore no hyperfine splitting. The results are shown in Table 1.

From the measured ratio $A^m/A = -1.73(25)$, where A^m and A are the hyperfine magnetic dipole constants of the isomeric state and the ground state, respectively, we determine the magnetic dipole moment μ^m of the isomer according to the relation $\mu^m = \mu A^m I^m / (A I)$, where μ indicates the magnetic moment of the ground state and I^m and I are the spin values of the isomeric and ground states, respectively³⁰. Here we neglect the HFS anomalies^{31,32}, which are small for the d^2 , fd and f^2 electronic configurations of the Th^{2+} levels (see Fig. 1), and derive $\mu^m/\mu = -1.04(15)$. The nuclear magnetic moment μ of the ground state has been measured in two experiments^{33,34} and the most precise value, $\mu = 0.360(7)\mu_N$ (where μ_N is the nuclear magneton), was obtained from high-precision calculations¹⁵ and measurements³⁴ of the HFS of $^{229}\text{Th}^{3+}$. On the basis of this value, we derive the magnetic dipole moment of the isomeric state as $\mu^m = -0.37(6)\mu_N$. An estimate based on the leading Nilsson configuration has predicted^{18,35} $\mu^m = -0.076\mu_N$. The discrepancy between these two results indicates that the simplified Nilsson approach is insufficient to quantitatively characterize the isomer because it neglects factors^{20,22} such as the expected collective quadrupole–octupole coupling of the nuclear deformation.

The spectroscopic quadrupole moment of the isomeric state Q_s^m is determined by $Q_s^m = Q_s B^m / B$, where Q_s is the spectroscopic quadrupole moment of the ground state. We use only the constants obtained for the 20711_1 electronic state to derive B^m/B (see Table 1).

Table 1 | Hyperfine constants of $^{229}\text{Th}^{2+}$ and $^{229\text{m}}\text{Th}^{2+}$ for the electronic levels 63₂ and 20711₁

| Level (cm ⁻¹) | Nuclear ground state | | Nuclear isomeric state | |
|---------------------------|----------------------|---------|------------------------|----------------------|
| | A (MHz) | B (MHz) | A ^m (MHz) | B ^m (MHz) |
| 63 | 151(8) | 73(27) | -263(29) | 53(65) |
| 20,711 | 88(4) | 897(14) | -151(22) | 498(15) |

The systematic uncertainty constitutes about 50% of the total uncertainty of the measurements. The main contributions to the systematic uncertainty are from the uncertainty in the reference cavity length, the instability of the frequency of the first-excitation-step laser and the nonlinearity of the frequency tuning of the second-step laser.

The measured ratio of the spectroscopic quadrupole moments is $Q_s^{\text{m}}/Q_s = 0.555(19)$. Q_s has been measured in two independent experiments^{12,15,34} to be 3.15(3) eb and 3.11(6) eb (where eb stands for electron-barn, 1 eb = 1.6022×10^{-47} C m²). Using the weighted mean value of these measurements, the quadrupole moment of the isomer is $Q_s^{\text{m}} = 1.74(6)$ eb. The spectroscopic quadrupole moment is related to the intrinsic quadrupole moment Q_0 through³⁰ $Q_s = Q_0(3K^2 - I(I+1))/((I+1)(2I+3))$, resulting in $Q_0^{\text{m}} = 8.7(3)$ eb for the isomeric state and $Q_0 = 8.8(1)$ eb for the ground state. Both states are the band-heads of their rotational bands and therefore the projection of the nuclear spin on the symmetry axis K is equal to I . The intrinsic quadrupole moments of the two states are the same ($Q_0^{\text{m}}/Q_0 = 0.99(4)$) within the uncertainty. Therefore, the nuclear charge distribution has a similar prolate shape in both states. This is in good agreement with theoretical predictions^{21,22}.

To investigate the difference of the charge radius between the ground state and the isomeric nucleus, the isomeric shifts of the first and second excitation steps are derived from the centres of the HFSs, calculated by setting $A = B = 0$. The isomeric HFS is shifted to a frequency 0.29(3) GHz lower than that of $^{229}\text{Th}^{2+}$ for the transition 63₂ → 20711₁. The isotope shift of this line between $^{229}\text{Th}^{2+}$ and $^{232}\text{Th}^{2+}$ is 8.2(2) GHz (see Fig. 2). The analysis of the data from the second excitation step, 20711₁ → 29300₀, yields an isomeric shift of 0.21(5) GHz and an isotopic shift of 6.2(3) GHz. The average ratio of the isomer and isotope shifts for both transitions is 0.035(4). The isotopic shift for heavy ions is primarily determined by the field shift, with only small corrections arising from the mass shift³⁶, and is therefore directly related with the nuclear charge radius. The measured isotope shifts correspond to the difference in the mean-square charge radii²⁹ $\langle r_{232}^2 \rangle - \langle r_{229}^2 \rangle = 0.33(5)$ fm², with $\langle r_{229}^2 \rangle \approx (5.76 \text{ fm})^2$. Consequently, the difference in the mean-square radii of the isomeric and ground states in ^{229}Th is $\langle r_{229\text{m}}^2 \rangle - \langle r_{229}^2 \rangle = 0.012(2)$ fm².

Discussion

The low-energy transition between the nuclear ground state and the isomeric state in ^{229}Th creates a bridge between atomic and nuclear physics and offers new perspectives for fundamental physics as well as for technological progress in metrology. We have measured the nuclear moments of $^{229\text{m}}\text{Th}$ and the isomeric shifts of two $^{229\text{m}}\text{Th}^{2+}$ lines. The determination of these basic nuclear properties allows us to calculate the HFS of $^{229\text{m}}\text{Th}$ for any electronic transition when the frequencies of ^{229}Th and another isotope with known nuclear radius²⁹ are known. This makes it possible to apply the sensitive electron–nuclear double-resonance detection⁴ of the isomer in the search for nuclear laser excitation and in nuclear clocks operating with trapped ions. This is particularly important when the long radiative lifetime of the isomer makes it difficult to detect photons emitted from the isomer decay and constitutes an important step towards the development of a nuclear clock.

The nuclear moments measured here enable more precise analyses^{4,5,37,38} of the expected systematic uncertainties of ^{229}Th nuclear clocks. In a clock based on ^{229}Th -doped crystals^{37,38}, the nuclear quadrupole moment interacts with crystal-field gradients, which may lead to a substantial frequency shift. In the trapped-ion clock, field-induced shifts will make only minor contributions to the uncertainty budget. The ^{229}Th nuclear clock has been proposed as a particularly sensitive system to search for temporal variations³⁹ in the fine-structure constant α .

This is based on a model in which the small transition energy $E_{\text{is}} \approx 7.8$ eV appears as the result of the nearly perfect cancellation of a change in the Coulomb energy $\Delta E_C = E_C^{\text{m}} - E_C \approx -1$ MeV by opposite and nearly equal changes of the nuclear energy through the strong interaction. Such a cancellation would be very sensitive to the values of the coupling constants of the electromagnetic and strong forces. This model has been criticized because the transition is performed by an unpaired neutron, leaving the Coulomb energies of ^{229}Th and $^{229\text{m}}\text{Th}$ essentially equal⁴⁰. By treating the nucleus as a uniform, hard-edged, prolate ellipsoid, the change in Coulomb energy can be expressed⁴¹ in terms of quantities that have been measured here:

$$\Delta E_C = (-485 \text{ MeV})[(\langle r_{229\text{m}}^2 \rangle / \langle r_{229}^2 \rangle) - 1] + (11.6 \text{ MeV})[(Q_0^{\text{m}}/Q_0) - 1] = -0.29(43) \text{ MeV}$$

The uncertainty in ΔE_C is dominated by the contribution from the $\pm 4\%$ uncertainty in Q_0^{m}/Q_0 . Although this result is not sufficient to prove that $|\Delta E_C| \gg E_{\text{is}}$, it is possible that the α -sensitivity of the ^{229}Th nuclear clock exceeds those of existing atomic clocks by several orders of magnitude. Combined with the expected high accuracy of the nuclear clock, this high sensitivity to α will enable us to test predictions of temporal variations of coupling constants and to experimentally assess theories unifying gravity with other interactions.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0011-8>.

Received: 15 September 2017; Accepted: 13 February 2018;

Published online 18 April 2018.

- Matinyan, S. Lasers as a bridge between atomic and nuclear physics. *Phys. Rep.* **298**, 199–249 (1998).
- Tkalya, E. V. Properties of the optical transition in the ^{229}Th nucleus. *Phys. Usp.* **46**, 315–320 (2003).
- Karpeshin, F. F. & Trzhaskovskaya, M. B. Impact of the electron environment on the lifetime of the $^{229}\text{Th}^{\text{m}}$ low-lying isomer. *Phys. Rev. C* **76**, 054313 (2007).
- Peik, E. & Tamm, C. Nuclear laser spectroscopy of the 3.5 eV transition in ^{229}Th . *Europhys. Lett.* **61**, 181–186 (2003).
- Campbell, C. J. et al. Single-ion nuclear clock for metrology at the 19th decimal place. *Phys. Rev. Lett.* **108**, 120802 (2012).
- Peik, E. & Okhapkin, M. Nuclear clocks based on resonant excitation of γ -transitions. *C. R. Phys.* **16**, 516–523 (2015).
- Kroger, L. A. & Reich, C. W. Features of the low energy level scheme of ^{229}Th as observed in the α -decay of ^{233}U . *Nucl. Phys. A* **259**, 29–60 (1976).
- Helmer, R. G. & Reich, C. W. An excited state of ^{229}Th at 3.5 eV. *Phys. Rev. C* **49**, 1845–1858 (1994).
- Barci, V., Ardisson, G., Barci-Funel, G., Weiss, B. & El Samad, O. Nuclear structure of ^{229}Th from γ spectroscopy study of ^{233}U α -particle decay. *Phys. Rev. C* **68**, 034329 (2003).
- Gulda, K. et al. The nuclear structure of ^{229}Th . *Nucl. Phys. A* **703**, 45–69 (2002).
- Ruchowska, E. et al. Nuclear structure of ^{229}Th . *Phys. Rev. C* **73**, 044326 (2006).
- Bemis, C. E. et al. Coulomb excitation of states in ^{229}Th . *Phys. Scr.* **38**, 657–663 (1988).
- Burke, D. G., Garrett, P. E., Qu, T. & Naumann, R. A. Nuclear structure of $^{229,231}\text{Th}$ studied with the $^{230,232}\text{Th}(d,t)$ reactions. *Nucl. Phys. A* **809**, 129–170 (2008).
- Nilsson, S. G. Binding states of individual nucleons in strongly deformed nuclei. *K. Dan. Vidensk. Selsk. Mat. Fys. Medd.* **29**, 1–69 (1955).
- Safronova, M. S., Safronova, U. I., Radnaev, A. G., Campbell, C. J. & Kuzmich, A. Magnetic dipole and electric quadrupole moments of the ^{229}Th nucleus. *Phys. Rev. A* **88**, 060501 (2013).
- Beck, B. R. et al. Energy splitting of the ground-state doublet in the nucleus ^{229}Th . *Phys. Rev. Lett.* **98**, 142501 (2007).
- Beck, B. R. et al. Improved Value for the Energy Splitting of the Ground-State Doublet in the Nucleus $^{229\text{m}}\text{Th}$. Internal Report LLNL-PROC-415170 (Lawrence Livermore National Laboratory, 2009); available at <https://e-reports-ext.llnl.gov/pdf/375773.pdf>.
- Dykhne, A. M. & Tkalya, E. V. Matrix element of the anomalously low-energy (3.5 ± 0.5 eV) transition in ^{229}Th and the isomer lifetime. *JETP Lett.* **67**, 251–256 (1998).
- Tkalya, E. V., Schneider, C., Jeet, J. & Hudson, E. R. Radiative lifetime and energy of the low-energy isomeric level in ^{229}Th . *Phys. Rev. C* **92**, 054324 (2015).
- Minkov, N. & Pálffy, A. Reduced transition probabilities for the gamma decay of the 7.8 eV isomer in ^{229}Th . *Phys. Rev. Lett.* **118**, 212501 (2017).

21. Tkalya, E. V. Proposal for a nuclear gamma-ray laser of optical range. *Phys. Rev. Lett.* **106**, 162501 (2011).
22. Litvinova, E., Feldmaier, H., Dobaczewski, J. & Flambaum, V. Nuclear structure of lowest ^{229}Th states and time-dependent fundamental constants. *Phys. Rev. C* **79**, 064303 (2009).
23. Jeet, J. et al. Results of a direct search using synchrotron radiation for the low-energy ^{229}Th nuclear isomeric transition. *Phys. Rev. Lett.* **114**, 253001 (2015).
24. Yamaguchi, A. et al. Experimental search for the low-energy nuclear transition in ^{229}Th with undulator radiation. *New J. Phys.* **17**, 053053 (2015).
25. von der Wense, L. et al. Direct detection of the ^{229}Th nuclear clock transition. *Nature* **533**, 47–51 (2016).
26. Seiferle, B., von der Wense, L. & Thierolf, P. G. Lifetime measurement of the ^{229}Th nuclear isomer. *Phys. Rev. Lett.* **118**, 042501 (2017).
27. Herrera-Sancho, O. A. et al. Two-photon laser excitation of trapped $^{232}\text{Th}^+$ ions via the 402-nm resonance line. *Phys. Rev. A* **85**, 033402 (2012).
28. Bjorkholm, J. E. & Liao, P. F. Line shape and strength of two-photon absorption in an atomic vapor with a resonant or nearly resonant intermediate state. *Phys. Rev. A* **14**, 751–760 (1976).
29. Kälber, W. et al. Nuclear radii of thorium isotopes from laser spectroscopy of stored ions. *Z. Phys. A* **334**, 103–108 (1989).
30. Kopfermann, H. *Nuclear Moments* (Academic, New York, 1958).
31. Bohr, A. & Weisskopf, V. F. The influence of nuclear structure on the hyperfine structure of heavy elements. *J. Phys. G* **77**, 94–98 (1950).
32. Büttgenbach, S. Magnetic hyperfine anomalies. *Hyperfine Interact.* **20**, 1–64 (1984).
33. Gerstenkorn, S. et al. Structures hyperfines du spectre d'étincelle, moment magnétique et quadrupolaire de l'isotope 229 du thorium. *J. Phys. France* **35**, 483–495 (1974).
34. Campbell, C. J., Radnaev, A. G. & Kuzmich, A. Wigner crystals of ^{229}Th for optical excitation of the nuclear isomer. *Phys. Rev. Lett.* **106**, 223001 (2011).
35. Chasman, R. R., Ahmad, I., Friedman, A. M. & Erskine, J. R. Survey of single-particle states in the mass region $A > 228^*$. *Rev. Mod. Phys.* **49**, 833–891 (1977).
36. King, W. H. *Isotope Shifts in Atomic Spectra* (Plenum, New York, 1984).
37. Kazakov, G. A. et al. Performance of a ^{229}Th solid-state nuclear clock. *New J. Phys.* **14**, 083019 (2012).
38. Rellergert, W. G. et al. Constraining the evolution of the fundamental constants with a solid-state optical frequency reference based on the ^{229}Th nucleus. *Phys. Rev. Lett.* **104**, 200802 (2010).
39. Flambaum, V. V. Enhanced effect of temporal variation of the fine structure constant and the strong interaction in ^{229}Th . *Phys. Rev. Lett.* **97**, 092502 (2006).
40. Hayes, A. C., Friar, J. L. & Möller, P. Splitting sensitivity of the ground and 7.6 eV isomeric states of ^{229}Th . *Phys. Rev. C* **78**, 024311 (2008).
41. Berengut, J. C., Dzuba, V. A., Flambaum, V. V. & Porsev, S. G. Proposed experimental method to determine α sensitivity of splitting between ground and 7.6 eV isomeric state in ^{229}Th . *Phys. Rev. Lett.* **102**, 210801 (2009).

Acknowledgements We thank C. Mokry, J. Runke, K. Eberhardt and N. G. Trautmann for the production of the ^{233}U source and M. Ehlers, S. Hennig and K. Kossert for the PTB ^{229}Th source. We acknowledge discussions with C. Tamm and B. Lipphardt and thank T. Leder, M. Menzel and A. Hoppmann for technical support. We acknowledge financial support from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement number 664732 (nuClock), from DFG through CRC 1227 (DQ-mat, project B04) and TH956-3-2 and from the LMU Department of Medical Physics via the Maier-Leibnitz Laboratory.

Reviewer information Nature thanks E. Hudson, M. Safronova and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.T. and M.V.O. developed the spectroscopy lasers. J.T., P.G., M.V.O., L.v.d.W., B.S., D.M.M. and P.G.T. did preparatory experimental work and performed the spectroscopy experiment. J.T., P.G., M.V.O. and E.P. performed the data analysis. M.V.O., P.G.T. and E.P. supervised the experiment. The ^{233}U source was produced by the group of C.E.D. All authors discussed the results. M.V.O., J.T., P.G. and E.P. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0011-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0011-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to E.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Ion trapping. *PTB trap.* Using laser ablation with a neodymium-doped yttrium–aluminium garnet (Nd:YAG) laser emitting 5-ns pulses with an energy of ≤ 1 mJ at 1,064 nm, the radio-frequency linear Paul trap located at PTB²⁷ is loaded with about 10^5 Th⁺ ions from a Th(NO₃)₃ solution (containing approximately equal amounts of ²²⁹Th and ²³²Th) dried on a tungsten target. Doubly charged thorium ions are generated in the trap via three-photon resonant ionization of Th⁺. For the first ionization step, a 402-nm ECDL with an output power of 20 mW, shaped into 90-ns-long pulses by an acousto-optical modulator, is used to pump the $24,874\text{ cm}^{-1}$ Th⁺ state. The second- and third-ionization-step photons are provided by third-harmonic generation of a pulsed titanium sapphire (Ti:Sa) laser (pulse length of 20 ns, third-harmonic generation peak power of about 1 kW) via the $63,258\text{ cm}^{-1}$ Th⁺ state⁴². Both lasers operate at a repetition rate of 1 kHz, and the pulses of the Ti:Sa laser and the ECDL are overlapped in time. The continuous resonant ionization produces a cloud of about $1,000\text{ }^{229}\text{Th}^{2+}$ ions and compensates the loss of ²²⁹Th²⁺ ions due to the formation of molecules with impurities of the buffer gas. By using argon as a buffer gas at a pressure of 0.1 Pa, we cool the ions to room temperature and depopulate metastable states by collisional quenching. Argon is used because it achieves a higher trap-loading efficiency than He. For the photodissociation of Th⁺ compounds formed in the trap⁴² we use the fourth-harmonic radiation of a Q-switched Nd:YAG laser operating at 266 nm with a pulse energy of about 10 μ J (see Extended Data Fig. 2 for the scheme of the optical setup). Owing to the presence of ²²⁹Th and ²³²Th isotopes in the solution on the target, we trap both isotopes simultaneously.

LMU trap. The radio-frequency quadrupole ion trap located at LMU is loaded with α -recoil thorium ions from a ²³³U source (source 3 in a previous publication²⁵). In the α decay of ²³³U, the low-energy isomeric state of ²²⁹Th is populated through a decay branch with an estimated branching ratio of 2%. Therefore, the trapped ion cloud consists of a mixture of ions in the ground and isomeric states. The ²³³U source was produced at the Institute for Nuclear Chemistry of the University of Mainz by molecular plating⁴³ (K. Eberhardt et al., manuscript in preparation). 290 kBq of ²³³U were deposited with a 90 mm diameter onto a Ti-sputtered Si wafer of 100 mm outer diameter and 0.5 mm thickness. The material contains ²³²U with a number fraction lower than 10^{-6} and was chemically purified by ion-exchange chromatography before deposition (25 months before the experiment) to remove the daughter isotopes of ²³³U and ²³²U. For the laser-spectroscopy experiments, the ²³³U source has a hole of 8 mm diameter⁴⁴ at the centre.

Because the kinetic energy of ²²⁹Th α -recoil ions is about 84 keV, about 10^5 recoils leave the ²³³U source per second. They are stopped in a buffer-gas stopping cell filled with 3.2×10^3 Pa of ultra-pure He 6.0, which is further purified by catalytic purification and a cryotrap^{44,45}. During the stopping process, charge exchange occurs between the thorium recoils and the buffer gas, producing predominantly thorium ions in the 2+ and 3+ charge states. After thermalization, these ions (together with the α -decay daughter products of the ²³³U decay chain) are guided by an electric radio-frequency and direct-current funnel system consisting of 50 ring electrodes towards a Laval extraction nozzle with a 0.6-mm-diameter nozzle throat. The extraction nozzle forms a supersonic gas jet and directs the ions into the subsequent (12-fold segmented) radio-frequency quadrupole, operated in this experiment as an ion trap. The extraction efficiency of the buffer gas stopping cell is about 5% and 10% in the 2+ and 3+ charge states, respectively, and thus more than 10,000 ²²⁹Th ions enter the radio-frequency quadrupole per second. In this way the trap is continuously loaded with ²²⁹Th ions; however, only about 1,000 ²²⁹Th²⁺ ions are actually trapped because of the trap's limited loading capacity. Besides ²²⁹Th, other isotopes originating from the α decay of nuclides other than ²³³U (for example, from ²³²U and its decay chain) or from sputtering can potentially enter the ion trap. These are listed in Extended Data Table 2, together with an estimation of their relative abundances compared to ²²⁹Th. Most of these isotopes are not potential background sources, because their relative abundances are too small. Isotopes that could affect the experiment are discussed in the sections 'Exclusion of ²³⁰Th isotope' and 'Exclusion of coincident absorption lines'. Owing to charge capture, the ²²⁹Th³⁺ ions are reduced to ²²⁹Th²⁺ after a few seconds. The He pressure in the trapping region is reduced to about 0.1 Pa by a differential pumping stage. To avoid the accumulation of molecular compounds in the trap, which are formed by chemical reactions with impurities, we empty the radio-frequency trap every 75 s. A new detection cycle starts 15 s after the trap reset, when the number of ²²⁹Th²⁺ ions in the trap reaches its maximum value of about 10^3 ions.

Spectroscopic lasers. The spectroscopy of the HFS of the thorium isomer is performed using continuous-wave ECDLs with a typical linewidth of 100 kHz. The ECDL at 1,164 nm has an output power of 30 mW and a tuning range greater than about 4 GHz. The blue ECDLs at 459 nm and 484 nm provide an output power of about 15 mW and a tuning range greater than 15 GHz. The radiation of the lasers is delivered to the trap by single-mode polarization-maintaining fibres. The power

of all lasers in both traps is about 4 mW owing to losses in fibre coupling and clipping by the supersonic nozzle, which corresponds to an intensity of 1.5 W cm^{-2} for each beam. The scheme of the optical setup is shown in Extended Data Fig. 2.

Laser frequency measurement and stabilization. To avoid long-term frequency drifts of the 459-nm laser and to provide controlled frequency steps of the 484-nm laser with an accuracy of a few megahertz, their wavelengths are stabilized to a Fizeau wavemeter (HighFinesse WS7) by a computer-based locking system. A Rb-stabilized ECDL at 780 nm is used to calibrate the wavemeter in intervals of 1,000 s. The 780-nm ECDL is stabilized to the $^2\text{S}_{1/2}(F=3) \rightarrow ^2\text{P}_{3/2}(F=4)$ ⁸⁵Rb line, where F is the total angular momentum, by the modulation-transfer spectroscopy technique⁴⁶. To measure the frequency detuning of the 1,164-nm laser during the scanning, we use a temperature-stabilized confocal cavity placed in vacuum.

Fluorescence detection. The fluorescence of the excited Th²⁺ ions is detected using a photomultiplier tube with a bandpass interference filter that transmits in the range of 445 ± 23 nm, which corresponds to one of the decay channels of the excited state and provides a background-free detection for the two-step excitation. A second photomultiplier tube is used for the measurements of the isotopic shift or, alternatively, to provide measurements of the number of thorium ions in the trap during the two-step excitation. This photomultiplier is set to detect the decay channels of the 484-nm or 459-nm excitations by using a filter that transmits at 643 ± 10 nm or 540 ± 8 nm, respectively, which blocks the laser stray light. Counters are used to register the photomultiplier signals. In the PTB trap, photon counting is terminated during the ionization and dissociation laser pulses to prevent them influencing the spectroscopic signal.

HFS. If a nucleus has spin $I > 1/2$, it may have a magnetic dipole moment and an electric quadrupole moment. The interaction of the valence electrons with these moments cause hyperfine splitting of the electronic levels, where the energy shift of an individual level is determined by

$$E_{\text{HFS}}(J, I, F) = \frac{AK}{2} + \frac{B[(3/4)K(K+1) - I(I+1)J(J+1)]}{2I(2I-1)J(2J-1)}$$

with $K = F(F+1) - J(J+1) - I(I+1)$. The hyperfine constants A and B are determined from the magnetic dipole and electric quadrupole interactions^{30,47}. For ²²⁹Th, the ground state has spin $I = 5/2$ and the isomer has $I^m = 3/2$. This leads to a splitting of the 63₂ electronic level into five sub-levels for the ground state and four sub-levels for the isomeric state. The 20711₁ level consists of three hyperfine levels in both nuclear states. Following the selection rules for electric dipole transitions ($\Delta F = \{0, \pm 1\}$), the spectrum of the two-step excitation with angular momentum $2 \rightarrow 1 \rightarrow 0$ for the ground and isomeric states consists of nine and eight resonances, respectively (see Extended Data Fig. 1).

Two-step excitation resonances. Assume laser 1 has a frequency detuning Δf_1 with respect to the centre of the first-step HFS. Δf_{gi} and Δf_{ie} are the frequency shifts of the individual hyperfine components of the first (gi) and the second (ie) step from the centres of their HFSs. The velocity class (v) of particles of an HFS component that are excited to the intermediate state is described as $k_1 v = -\Delta f_{\text{gi}} + \Delta f_1$. Particles with the same v are excited to the upper state according to the equation $\pm(k_2 v) = -\Delta f_{\text{ie}} + \Delta f_2$, where Δf_2 is the second laser frequency detuning from the second-step HFS centre, $+k_2 v$ ($-k_2 v$) corresponds to co- (counter-) propagating beams and $k_{1,2}$ are the wavevectors of the first and second excitation steps. By combining the two excitation steps for ions with the same v , the position of an individual narrow two-step resonance (free from Doppler broadening) is represented as $\Delta f_2 = \pm(k_2/k_1)(-\Delta f_{\text{gi}} + \Delta f_1) + \Delta f_{\text{ie}}$. The amplitude of the resonance depends on the fraction of ions within the velocity group of the Doppler distribution, which interacts with the first-step laser radiation, and the product of $\Omega_{2\text{gi}}$ and $\Omega_{2\text{ie}}$, where Ω are the Rabi frequencies of the transitions of the two excitation steps²⁸.

HFS constants. The intervals between the resonances are measured with respect to the transmission peaks of the stable confocal cavity, which are recorded simultaneously with the HFS resonances. The measured frequency intervals of the HFS are fitted with a least-squares fit according to the equations described in the previous sections to determine A and B for the 63₂ and the 20711₁ states. The algorithm attempts fits for all viable assignment combinations of F values for the resolved transitions. The isomeric state is fitted without requiring fixed ratios A^m/A and B^m/B for both electronic states.

Map of the HFS resonances. Extended Data Fig. 3 shows selected spectra demonstrating the evolution of the HFS peaks of the ground and isomeric states for different frequency positions of the first excitation step. The isomeric-state HFS resonances are marked with red labels and all resonances are described in Extended Data Table 1. The evaluation of the positions and the amplitudes of the seven additional peaks that are only observed in the LMU trap over the mapping indicates that those resonances fit to the HFS pattern of an $I = 3/2$ isomer. We tested for the appearance of spurious resonances from the collision-induced intermediate-state

changes, possible laser multimode operation and back-reflection of the spectroscopic laser beams (mixing of co- and counter-propagating beam geometries) and confirmed that the observed resonances are genuine spectroscopic features.

Extended Data Fig. 4 was generated by tracking the positions and amplitudes of the second-step resonances for all acquired spectra in both nuclear states. The figure shows the relative amplitudes of the resonances and was used to calculate the fraction of ions in the isomeric state.

Exclusion of ^{230}Th isotope. Estimations for daughter products from the ^{233}U source (Extended Data Table 2) show upper limits for the presence of the ^{228}Th and ^{230}Th isotopes. For all transitions, the resonance of ^{230}Th ($I = 0$, no HFS) should appear between the lines of ^{232}Th and ^{229}Th , like those of the HFS of $^{229\text{m}}\text{Th}$. The isotopic shift of $^{230}\text{Th}^{2+}$ with respect to $^{229}\text{Th}^{2+}$ for the 484-nm transition is 3.2 GHz, as calculated from previous measurements at 459 nm using the LMU trap with a ^{234}U source and determined in an earlier work²⁹. This shift is outside the range of the observed isomeric HFS. In the experiment with the ^{233}U source, the $^{230}\text{Th}^{2+}$ resonance is not observed, indicating a considerably smaller flux than that listed in Extended Data Table 2.

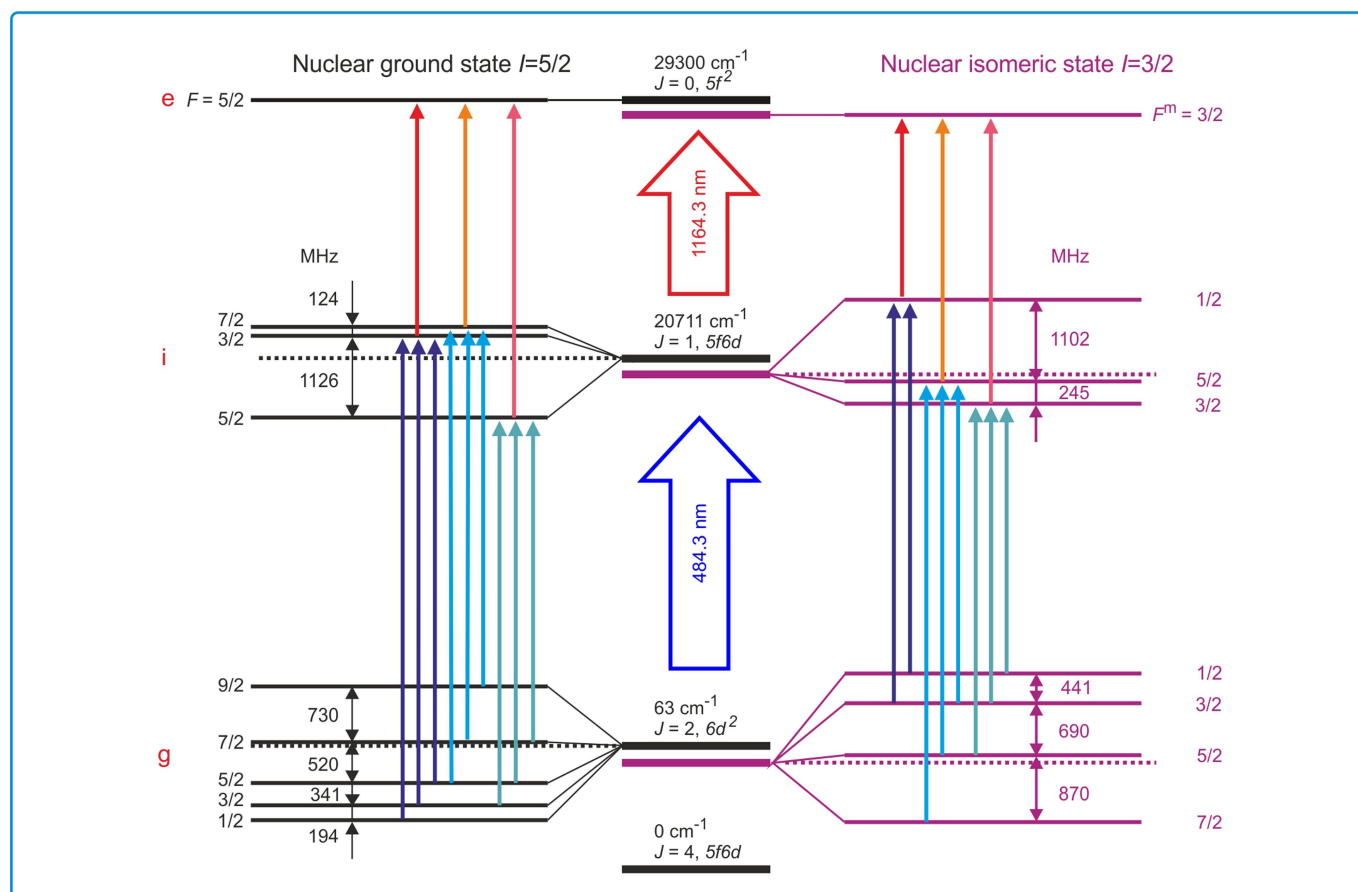
Exclusion of coincident absorption lines. Because the ^{233}U source emits a variety of ions of different elements (α -decay daughter products of the ^{233}U decay chain), which are loaded into the trap simultaneously with the thorium ions (see Extended Data Table 2), spectral lines of those elements might be detected in parallel to the isomer signal. Only two elements (U and Pu) have a flux high enough to be detected. For the first excitation step, the uranium and plutonium lines that are closest to the $^{229}\text{Th}^{2+}$ resonances and that originate from low-lying levels are detuned by about 200 GHz ($^{238}\text{U}^{+}$, $\bar{\nu} = 20,640.51\text{ cm}^{-1}$, transition from $915_{9/2}$ to $21555_{9/2}$) and about 60 GHz ($^{240}\text{Pu}^{+}$, $\bar{\nu} = 20,645.62\text{ cm}^{-1}$, transition from $3970_{5/2}$ to $24615_{3/2}$)^{48–50}. We exclude also the influence of Th^{2+}He complexes, which can be formed by interaction with the buffer gas: owing to their low binding energy, estimated⁵¹ to be less than $7,500\text{ cm}^{-1}$, these complexes would be dissociated by laser excitation at 484 nm ($20,648\text{ cm}^{-1}$). Moreover, because the signal that we detect requires coincidence of two resonance conditions, we can rule out that the recorded thorium spectra are affected by other elements.

Isomeric lifetime in Th^{2+} . It is in principle possible to determine the isomeric lifetime in Th^{2+} by measuring the time evolution of the amplitudes of the resonances for both nuclear states. The isomeric signal will show an additional exponential decay due to its finite lifetime. This experiment is at present limited by the storage time of the ions in the trap (about 60 s), which is defined by chemical reactions and charge exchange with impurities in the buffer gas. Therefore, this value can only be given as a lower limit for the isomer lifetime. By improving the storage time by two orders of magnitude, measuring the isomeric lifetime will become feasible and provide an important parameter of the clock.

Sensitivity to the fine-structure constant. The ^{229}Th nuclear clock has been proposed³⁹ as a particularly sensitive system to search for temporal variations of the fine structure constant α , but this proposal has been met with scepticism⁴⁰. The sensitivity $(\alpha/f)(df/d\alpha)$ of the nuclear transition frequency f to the value of α is equal to the ratio $\Delta E_{\text{C}}/E_{\text{is}}$ of the change in Coulomb energy (ΔE_{C}) and the total transition energy ($E_{\text{is}} \approx 7.8\text{ eV}$) between the ground and isomeric states. Theoretical estimations of ΔE_{C} from nuclear structure calculations^{22,40} vary from a few kiloelectronvolts to a few megaelectronvolts. It has been proposed⁴¹ that ΔE_{C} can be calculated via the change of the nuclear charge radius and electric quadrupole moment from measured isomer shift and HFS data. Applied to our data, using updated¹⁵ values of Q_0 and $\langle r_{229}^2 \rangle$, this results in $\Delta E_{\text{C}} = -0.29(43)\text{ MeV}$. Because the change in the charge radius $\Delta\langle r^2 \rangle/\langle r^2 \rangle \approx 4 \times 10^{-4}$ is small, the uncertainty in the change of the quadrupole moments, which is known from $Q_0^{\text{m}}/Q_0 = 0.99(4)$, yields the dominant uncertainty contribution to ΔE_{C} . Although smaller values cannot be excluded with certainty, the most probable modulus for the α -sensitivity of a ^{229}Th nuclear clock is about 4×10^4 .

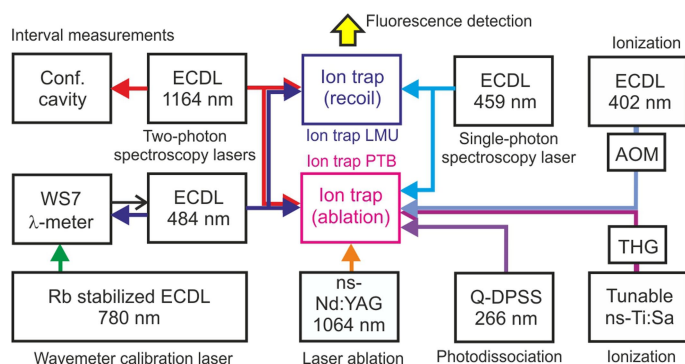
Data availability. Source Data for Figs. 2–4 and Extended Data Figs. 3–5 are provided in the online version of the paper. Further data are available at <https://zenodo.org/communities/nuclock/> and from the corresponding author upon reasonable request.

42. Herrera-Sancho, O. A., Nemitz, N., Okhapkin, M. V. & Peik, E. Energy levels of Th^{+} between 7.3 and 8.3 eV. *Phys. Rev. A* **88**, 012512 (2013).
43. Vascon, A. et al. Elucidation of constant current density molecular plating. *Nucl. Instrum. Meth. A* **696**, 180–191 (2012).
44. von der Wense, L. *On the Direct Detection of $^{229\text{m}}\text{Th}$* . PhD thesis, Ludwig-Maximilians-Universität, Munich (2016).
45. von der Wense, L., Seiferle, B., Laatiaoui, M. & Thierolf, P. G. Determination of the extraction efficiency for ^{233}U source α -recoil ions from the MLL buffer-gas stopping cell. *Eur. Phys. J. A* **51**, 29 (2015).
46. McCarron, D. J., King, S. A. & Cornish, S. L. Modulation transfer spectroscopy in atomic rubidium. *Meas. Sci. Technol.* **19**, 105601 (2008).
47. Sobelman, I. I. *Atomic Spectra and Radiative Transitions* (Springer, Berlin, 1979).
48. Palmer, B. A., Keller, R. A. & Engleman Jr., R. *An Atlas of Uranium Emission Intensities in a Hollow Cathode Discharge*. Report No. LA-8251-MS (Los Alamos Scientific Laboratory, 1980).
49. Blaise, J. & Wyart, J.-F. *Database of Selected Constants, Energy Levels and Atomic Spectra of Actinides* <http://web2.lac.u-psud.fr/lac/Database/Contents.html> (2017).
50. Blaise, J., Fred, M. & Gutmacher, R. G. *The Atomic Spectrum of Plutonium*. Report No. ANL-83-95 (Argonne National Laboratory, 1983).
51. Wright, T. G. & Breckenridge, W. H. Radii of atomic ions determined from diatomic ion–He bond lengths. *J. Phys. Chem. A* **114**, 3182–3189 (2010).

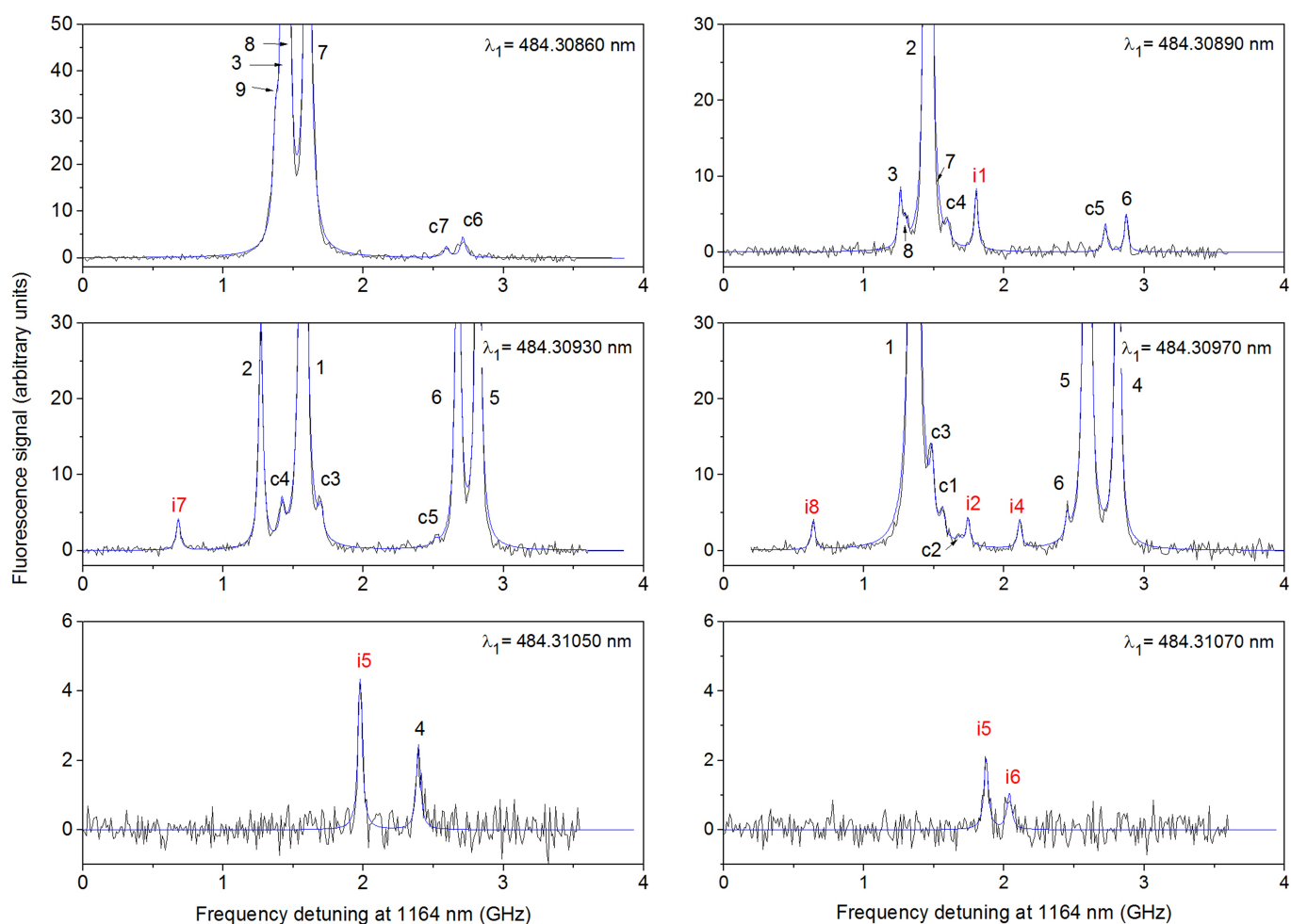


Extended Data Fig. 1 | Detailed level scheme of the two-step excitation. Transitions and electronic configurations of the initial (g), intermediate (i) and excited (e) states relevant to the experiment are shown, labelled by their energy in cm^{-1} and the electronic angular momentum J . Hyperfine

sub-levels are indicated by their total angular momentum F and F^m . Transitions belonging to the same intermediate hyperfine level are depicted with the same colour. The hyperfine intervals are calculated from the hyperfine constants A and B presented in Table 1.

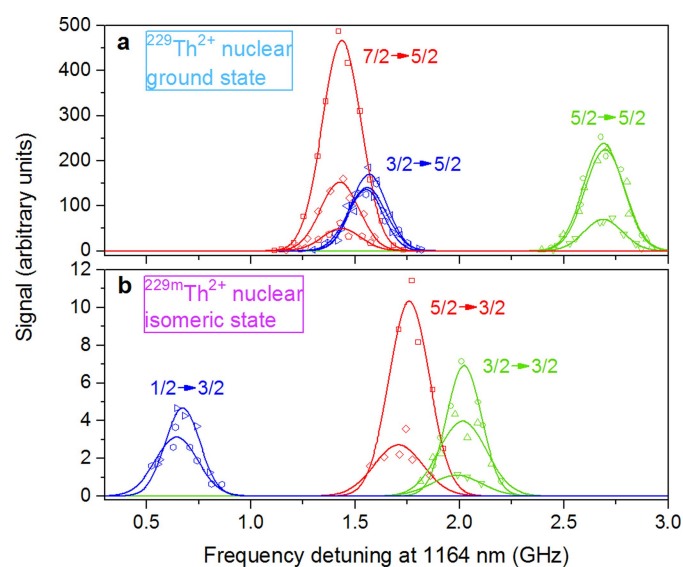


Extended Data Fig. 2 | Scheme of the optical setup. The spectroscopy laser of the first step excitation (484 nm) is locked to the wavemeter, which is calibrated by a Rb-stabilized ECDL at 780 nm. The second-step (1,164 nm) laser tuning is monitored with the confocal cavity. The ECDL at 459 nm is used to detect the number of ions in the traps. The loading of Th^{2+} in the PTB trap is provided by ablation (nanosecond Nd:YAG laser at 1,064 nm) and further three-photon ionization. The first step uses a 402-nm ECDL, pulsed via an acousto-optical modulator (AOM), and the second and third steps involve third-harmonic generation (THG) of a nanosecond Ti:Sa laser. Molecular compounds of Th^+ are photodissociated by pulses from a Q-switched diode-pumped solid-state laser (Q-DPSS).

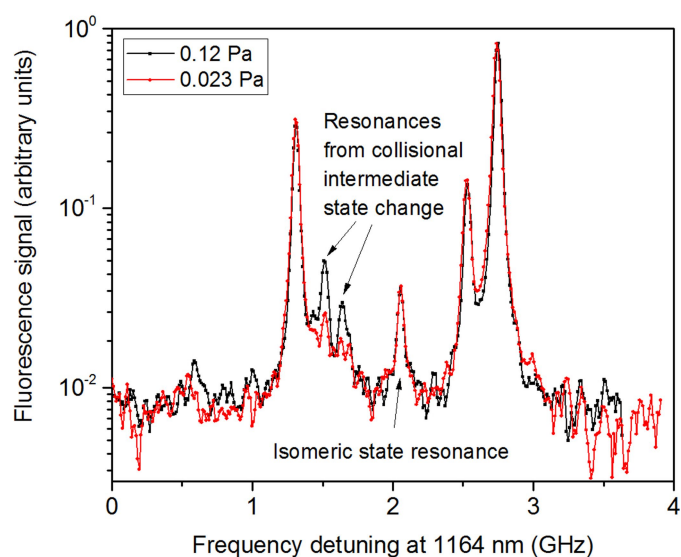


Extended Data Fig. 3 | Selected spectra obtained by two-step excitation. The resonances recorded for different positions of the 484-nm ECDL show the observed isomeric peaks for the case of co-propagating beams (labelled 'i'). The resonances that originate from collisions of ions in the intermediate state are labelled 'c'. The description of the peaks and

their total angular momenta are given in Extended Data Table 1. Black lines show the recorded data and blue lines represent a multi-Lorentz fit with fixed width, which is used to extract the line centres and frequency intervals.



Extended Data Fig. 4 | Mapping of the second excitation step. The experimental points represent amplitudes and positions of the two-step resonances obtained by setting the 484-nm laser at certain frequencies and tuning the 1,164-nm laser. The frequency of the 484-nm laser is changed in steps of about 120 MHz. The resonance groups shown with the same colour correspond to transitions from the same intermediate state with total angular momentum F , which is populated from different ground-state hyperfine components. The graphs show the HFS transitions of $^{229}\text{Th}^{2+}$ in the ground state (a) and the isomer (b).



Extended Data Fig. 5 | Pressure dependence of collision-induced changes in the intermediate-state HFS. The two-step excitation resonances of Th^{2+} were obtained with the first laser stabilized at -800 MHz detuning with respect to the ^{229}Th HFS centre and the second laser scanned. The measurement is performed for two different He buffer-gas pressures and shows a decrease in the relative amplitude of the collisional resonances for the reduction of the buffer-gas pressure. We note that the isomeric resonance is not affected by the change in He pressure.

Extended Data Table 1 | Systematics of the observed resonances

| Label | $F_g \rightarrow F_i \rightarrow F_e$ | Label | $F_g \rightarrow F_i; F_i \rightarrow F_e$ | Label | $F_g^m \rightarrow F_i^m \rightarrow F_e^m$ |
|-------|---------------------------------------|-------|--|-------|---|
| 1 | $9/2 \rightarrow 7/2 \rightarrow 5/2$ | c1 | $7/2 \rightarrow 5/2; 7/2 \rightarrow 5/2$ | i1 | $7/2 \rightarrow 5/2 \rightarrow 3/2$ |
| 2 | $7/2 \rightarrow 7/2 \rightarrow 5/2$ | c2 | $7/2 \rightarrow 5/2; 3/2 \rightarrow 5/2$ | i2 | $5/2 \rightarrow 5/2 \rightarrow 3/2$ |
| 3 | $5/2 \rightarrow 7/2 \rightarrow 5/2$ | c3 | $9/2 \rightarrow 7/2; 3/2 \rightarrow 5/2$ & | i3* | $3/2 \rightarrow 5/2 \rightarrow 3/2$ |
| 4 | $7/2 \rightarrow 5/2 \rightarrow 5/2$ | | $5/2 \rightarrow 5/2; 3/2 \rightarrow 5/2$ | i4 | $5/2 \rightarrow 3/2 \rightarrow 3/2$ |
| 5 | $5/2 \rightarrow 5/2 \rightarrow 5/2$ | c4 | $7/2 \rightarrow 7/2; 3/2 \rightarrow 5/2$ | i5 | $3/2 \rightarrow 3/2 \rightarrow 3/2$ |
| 6 | $3/2 \rightarrow 5/2 \rightarrow 5/2$ | c5 | $7/2 \rightarrow 7/2; 5/2 \rightarrow 5/2$ | i6 | $1/2 \rightarrow 3/2 \rightarrow 3/2$ |
| 7 | $5/2 \rightarrow 3/2 \rightarrow 5/2$ | c6 | $5/2 \rightarrow 3/2; 5/2 \rightarrow 5/2$ | i7 | $3/2 \rightarrow 1/2 \rightarrow 3/2$ |
| 8 | $3/2 \rightarrow 3/2 \rightarrow 5/2$ | c7 | $3/2 \rightarrow 3/2; 5/2 \rightarrow 5/2$ | i8 | $1/2 \rightarrow 1/2 \rightarrow 3/2$ |
| 9 | $1/2 \rightarrow 3/2 \rightarrow 5/2$ | | | | |

The detected resonances are listed with the total angular momenta of the electronic states involved in the excitation. The resonances of the nuclear ground state are labelled with numbers. The resonances that arise from collisional changes of the intermediate state population are described by both quantum numbers, F_i (before the collision) and F_i' (after the collision), and are labelled 'c'. Isomeric resonances are marked with 'i'. The resonance i3 (marked with an asterisk) is not observed in the experiment.

Extended Data Table 2 | Extraction of isotopes from the ^{233}U source

| Isotope | Extraction (rel. to Th-229) | Isotope | Extraction (rel. to Th-229) |
|--------------------|--------------------------------|--------------------|--------------------------------|
| Th-229 | 1 | Pu-238 | 4×10^{-6} |
| U-233 | ≈ 1 | Pu-239 | 2×10^{-3} |
| Th-229 decay chain | 2×10^{-4} | U-235 | 1×10^{-2} |
| U-232 | 6×10^{-7} | Pu-240 | 4×10^{-4} |
| Th-228 | 1×10^{-3} | U-236 | 1×10^{-2} |
| Th-228 decay chain | 7×10^{-4} | Pa-231 | 2×10^{-5} |
| U-234 | 2×10^{-2} | Ac-227 | 1×10^{-4} |
| Th-230 | 7×10^{-3} | Ac-227 decay chain | 6×10^{-6} |

The numbers of recoils leaving the ^{233}U source via α decay and sputtering are listed. The extraction ratios show the upper limit on the number of recoils relative to the number of ^{229}Th recoils.

Functional circuit architecture underlying parental behaviour

Johannes Kohl¹, Benedicte M. Babayan², Nimrod D. Rubinstein¹, Anita E. Autry¹, Brenda Marin-Rodriguez¹, Vikrant Kapoor², Kazunari Miyamishi³, Larry S. Zweifel^{4,5}, Liqun Luo³, Naoshige Uchida² & Catherine Dulac^{1*}

Parenting is essential for the survival and wellbeing of mammalian offspring. However, we lack a circuit-level understanding of how distinct components of this behaviour are coordinated. Here we investigate how galanin-expressing neurons in the medial preoptic area (MPOA^{Gal}) of the hypothalamus coordinate motor, motivational, hormonal and social aspects of parenting in mice. These neurons integrate inputs from a large number of brain areas and the activation of these inputs depends on the animal's sex and reproductive state. Subsets of MPOA^{Gal} neurons form discrete pools that are defined by their projection sites. While the MPOA^{Gal} population is active during all episodes of parental behaviour, individual pools are tuned to characteristic aspects of parenting. Optogenetic manipulation of MPOA^{Gal} projections mirrors this specificity, affecting discrete parenting components. This functional organization, reminiscent of the control of motor sequences by pools of spinal cord neurons, provides a new model for how discrete elements of a social behaviour are generated at the circuit level.

Although essential for survival at a multigenerational time scale, parental care entails sacrifices without immediate benefits for the caregiver, suggesting that this behaviour is driven by evolutionarily shaped, hard-wired neural circuits^{1,2}. Parenting, similar to other naturalistic behaviours, comprises multiple coordinated components, such as specific motor patterns, an enhanced motivation to interact with infants, distinct hormonal states and often the suppression of other social activities such as mating. We aimed to exploit the recent identification of MPOA^{Gal} neurons as a key node in the control of parenting in mice³ to uncover organizational principles of associated neural circuits. We hypothesized that the function of MPOA^{Gal} neurons in parental behaviour requires integration of external signals, such as stimuli from pups and other environmental sources, and internal hormonal and metabolic information, as well as the ability to coordinate the motor, motivational, hormonal and social components of parenting.

Identity and activity of MPOA^{Gal} inputs

To determine brain-wide inputs into MPOA^{Gal} neurons, we used rabies virus-mediated retrograde trans-synaptic tracing⁴ (Fig. 1a), and found that MPOA^{Gal} neurons receive direct inputs from more than 20 areas in both male and female mice (Fig. 1b, c, Extended Data Fig. 1a and Extended Data Table 1). Presynaptic neurons within the MPOA itself provided the highest fractional input (approximately 20%), and hypothalamic inputs accounted for about 60% of the presynaptic neurons, suggesting that extensive local processing occurs (Fig. 1c). MPOA^{Gal} neurons also receive inputs from monoaminergic and neuropeptidergic modulatory areas, the mesolimbic reward system, pathways associated with pheromone-processing, and hypothalamic as well as septal areas involved in emotional states (Fig. 1c and Extended Data Fig. 1a). Inputs from the paraventricular hypothalamic nucleus (PVN), a key area for homeostatic and neuroendocrine control, were particularly abundant. Notably, MPOA^{Gal} neurons did not receive direct inputs from oxytocin (OXT)-secreting PVN (PVN^{OXT}) neurons, which are implicated in parturition, lactation and maternal behaviour^{1,2,5}, but instead received inputs from vasopressin-expressing PVN (PVN^{AVP}) neurons, which are associated with the modulation of many social behaviours⁶ and nest

building⁷ (Fig. 1d). MPOA^{Gal} neurons also received inputs from AVP⁺, but not OXT⁺, neurons of the supraoptic nucleus (Extended Data Fig. 1d). Input fractions were similar in males and females, with a few exceptions (Fig. 1e, f and Extended Data Fig. 1a). Therefore, MPOA^{Gal} neurons appear to be anatomically well-positioned to integrate external (sensory) as well as internal (modulatory) signals that are relevant to parenting in both sexes.

Next, we investigated MPOA^{Gal} input activation during parenting according to the animal's sex and reproductive state. In laboratory mice, virgin females and sexually experienced males and females show parental behaviours, whereas virgin males typically attack and kill pups^{3,8,9}. We combined rabies tracing with immunostaining for the activity marker Fos after parenting in primiparous females (mothers), virgin females and fathers (Fig. 1g) and compared the Fos⁺ fraction of input neurons between parental animals and non-pup-exposed controls (Fig. 1h–j). Local MPOA inputs were specifically activated during parenting in all groups (Fig. 1h–j), whereas the activation of other inputs was dependent on sex and reproductive state: in parents, but not virgin females, a subset of reward-associated and modulatory inputs were activated (Fig. 1h–j). Presynaptic neurons in pheromone-processing pathways (the medial amygdala (MeA) and bed nucleus of the stria terminalis (BNST)) were selectively activated in fathers and virgin females, but not in mothers (Fig. 1h–j). Because pup-directed aggression in virgin mice is pheromone-dependent^{3,8}, the MeA–BNST pathway might remain partially active in sexually experienced males and parental virgin females, whereas it is fully silenced only in mothers. Intriguingly, the largest number of inputs was activated in fathers (Fig. 1j), and non-overlapping subsets of inputs were activated in mothers and virgin females (Fig. 1h, i). These results suggest that MPOA^{Gal} neurons perform different computations of inputs according to the animal's sex and reproductive state.

Input–output logic of the MPOA^{Gal} circuit

To identify MPOA^{Gal} projections and synaptic targets, we infected MPOA^{Gal} neurons with adeno-associated viruses (AAVs) encoding the fluorophore tdTomato as well as the presynaptic marker

¹Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, MA, USA. ²Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, MA, USA. ³Howard Hughes Medical Institute, Department of Biology, Stanford University, Stanford, CA, USA. ⁴Department of Pharmacology, University of Washington, Seattle, WA, USA. ⁵Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA. *e-mail: dulac@fas.harvard.edu

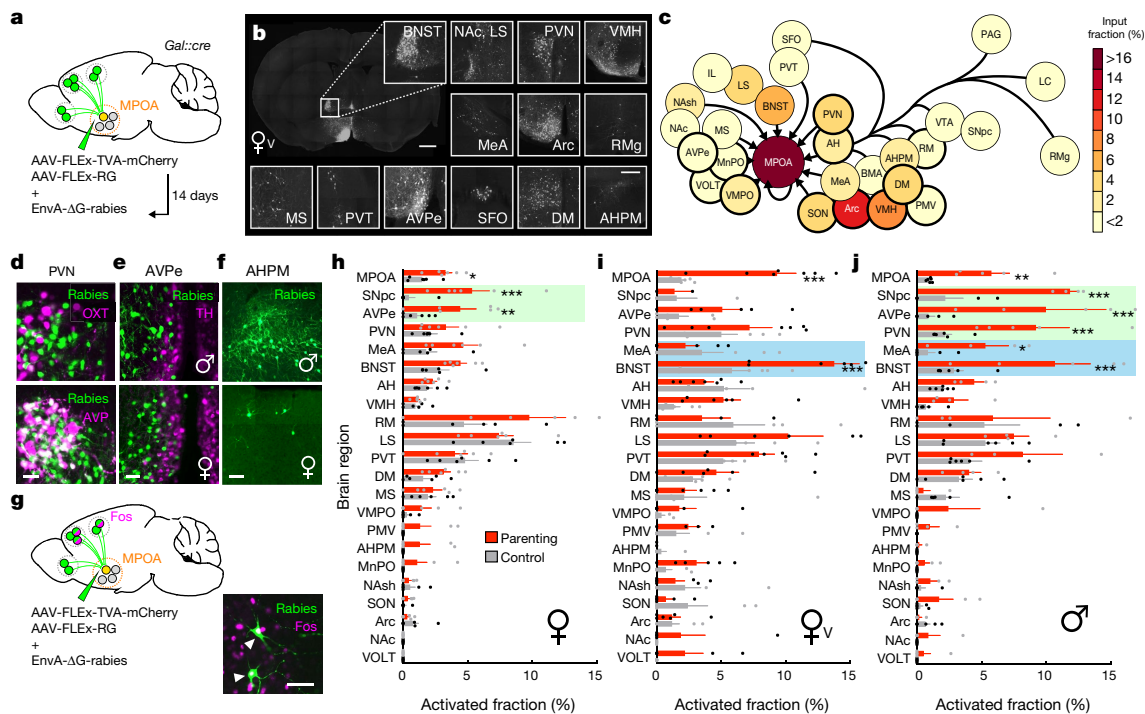


Fig. 1 | MPOA^{Gal} inputs are activated during parental behaviour in a sex- and reproductive state-specific manner. **a**, Monosynaptic retrograde tracing from MPOA^{Gal} neurons. **b**, Input areas with rabies⁺ neurons in a virgin female. **c**, Overview of inputs into MPOA^{Gal} neurons. Hypothalamic input areas are circled in bold. **d**, MPOA^{Gal} neurons receive monosynaptic inputs from magnocellular PVN^{AVP} (37.6 ± 4.1% overlap, *n* = 3 mice) but rarely from PVN^{OXT} (2.6 ± 0.6%, *n* = 3 mice) neurons. **e**, Presynaptic neurons in AVPe are TH⁺ in males (1.9% TH⁺, *n* = 2 mice) and females (1.8% TH⁺, *n* = 3 mice). **f**, Presynaptic neurons in posteromedial amygdalo-hippocampal area (AHPM). **g**, Identification of activated

MPOA^{Gal} inputs and example of Fos⁺ presynaptic neurons. **h–j**, Activated input fractions in mothers (**h**), virgin females (**i**) and fathers (**j**). *n* = 6 pup-exposed mice, *n* = 6 controls each. Green boxes, parent-specific activation; blue boxes, father- and virgin female-specific activation. Two-tailed *t*-tests (corrected for multiple comparisons, Methods); **h**, ****P* < 0.0001, ***P* = 0.0267, **P* = 0.0196; **i**, ****P* < 0.0001; **j**, ****P* < 0.0001, ***P* = 0.0035, **P* = 0.0104. **h–j**, Data are mean ± s.e.m.; *n* = number of mice in all figures. Scale bars, 500 μm (**b**, left), 250 μm (**b**, inset) and 50 μm (**d–g**). For definitions of the abbreviations, see Extended Data Table 1.

synaptophysin conjugated to GFP (Syn-GFP; Fig. 2a and Extended Data Fig. 2a). MPOA^{Gal} neurons project to approximately 20 areas in males and females (Fig. 2b, c and Extended Data Fig. 2b). Many of these regions were previously shown to be involved in maternal behaviour using pharmacological manipulations and lesions, mainly in rats¹⁰ (Extended Data Table 2). Notably, this projection map mostly overlaps with the input map defined above (Fig. 1c), revealing extensive reciprocal connectivity in parental circuits.

Among the areas most intensely labelled by Syn-GFP were the PVN and anteroventral periventricular nucleus (AVPe) (Fig. 2c), which have both been implicated in the control of parenting^{6,11}. Using rabies tracing from molecularly defined PVN cell types (Fig. 2d), we found that MPOA^{Gal} neurons project to PVN^{AVP}, PVN^{OXT} and corticotropin-releasing hormone (CRH)-expressing PVN neurons (PVN^{CRH}) in both males and females (Fig. 2e–g). Furthermore, connectivity from MPOA^{Gal} neurons to PVN neurons appears sexually dimorphic, with more MPOA^{Gal} neurons projecting to PVN^{AVP} and PVN^{CRH} neurons in males and more MPOA^{Gal} neurons projecting to PVN^{OXT} neurons in females (Fig. 2e–g). MPOA^{Gal} neurons might therefore exert control over parenting-promoting hormonal release in a sex-specific fashion.

Tyrosine-hydroxylase (TH)-expressing neurons in the AVPe were found to influence parenting in females via monosynaptic connections¹¹ from AVPeTH to PVN^{OXT} neurons. Rabies tracing from MPOA^{Gal} or AVPeTH neurons showed that whereas MPOA^{Gal} neurons do not receive monosynaptic inputs from AVPeTH neurons (Fig. 1e), AVPeTH neurons do receive direct inputs from MPOA^{Gal} neurons in both males and females (Extended Data Fig. 2e, f). Thus, MPOA^{Gal} neurons might also influence OXT secretion via a disynaptic circuit from MPOA^{Gal} → AVPeTH → PVN^{OXT} neurons (Extended Data Fig. 2g).

We next investigated the organization of MPOA^{Gal} projections, and their activity during parenting. Injections of the retrograde tracer

cholera toxin subunit B (CTB) into pairs of MPOA^{Gal} projection targets revealed few double-labelled MPOA^{Gal} neurons (Extended Data Fig. 3a–c). Moreover, retrogradely labelled cell bodies from individual projections occupied characteristic, mostly non-overlapping zones in the MPOA (Extended Data Fig. 3f, g) and conditional tracing of individual projection areas identified only minor collaterals (Extended Data Fig. 4). These results suggest that MPOA^{Gal} neurons are organized in distinct pools, each projecting to mostly non-overlapping target areas. To assess whether different MPOA^{Gal} pools, as defined by their projection sites, were equally activated during parenting, we used a Cre-dependent, retrograde canine adenovirus (CAV) to label MPOA^{Gal} subpopulations projecting to regions that have previously been implicated in parenting (12 out of 22 projections; Extended Data Table 2) and quantified their activation in parental females (Fig. 2h). Fractions of Fos⁺ neurons differed widely between projections, ranging from more than 50% (projections to the periaqueductal grey (PAG)) to less than 10% (projections to the ventromedial hypothalamus, Fig. 2i). A similar distribution was found in parental fathers (Extended Data Fig. 2d).

On the basis of their high projection density (Fig. 2c), high activity during parenting (Fig. 2i) and potentially diverse contributions to this behaviour (Extended Data Table 2), we selected MPOA^{Gal} subpopulations that projected to the PAG, MeA, ventral tegmental area (VTA) and PVN for further characterization. Gal⁺ neurons were approximately twice more likely to project to most of these candidate areas than expected from their frequency in the MPOA (Extended Data Fig. 3d, e), supporting the hypothesis that these projections have prominent roles in the control of parenting.

We next aimed to determine whether projection-defined MPOA^{Gal} subpopulations receive selected inputs from the approximately 20 identified upstream areas (Fig. 1c) or whether they uniformly integrate all inputs. We used a double-conditional approach in which rabies virus

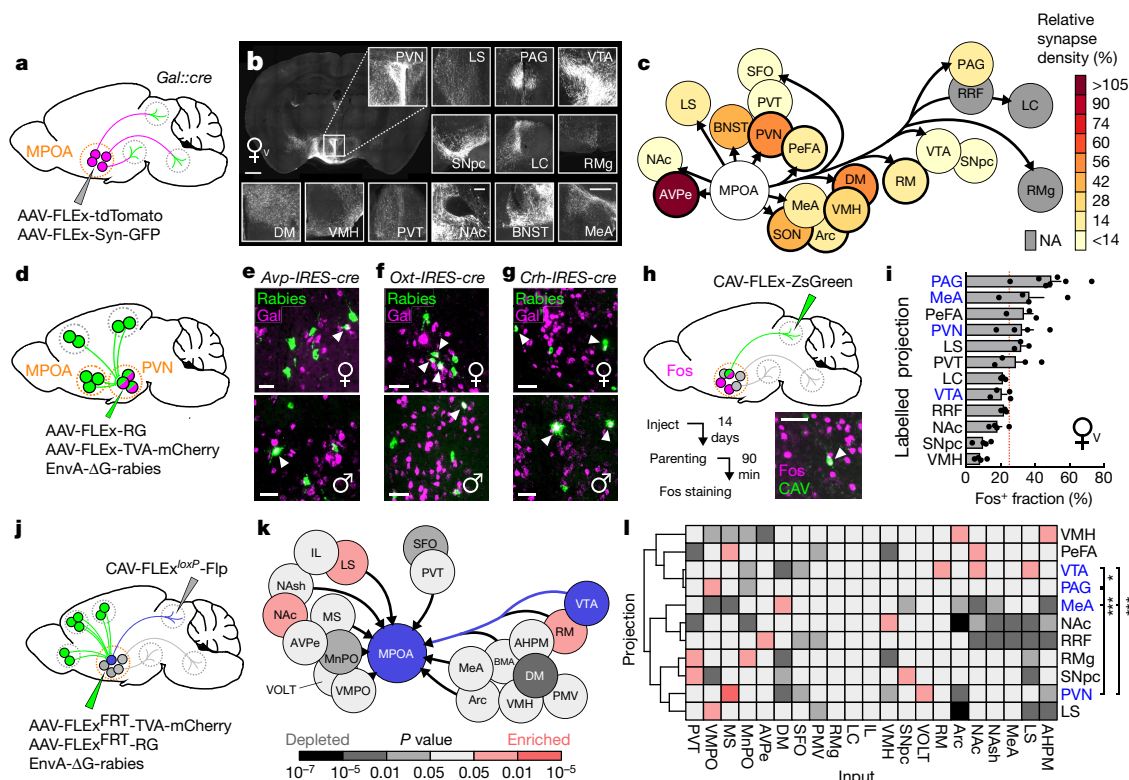


Fig. 2 | Identification of parenting-activated MPOA^{Gal} projections and input-output logic of the MPOA^{Gal} circuit. **a**, Visualization of MPOA^{Gal} projections. **b**, MPOA^{Gal} projections identified by tdTomato fluorescence in virgin females. **c**, Relative synaptic density in MPOA^{Gal} projection targets ($n = 4$ mice, Methods). Grey regions could not be quantified owing to tissue autofluorescence. Hypothalamic target areas are circled in bold. **d**, Monosynaptic retrograde tracing from PVN. **e–g**, MPOA^{Gal} neurons are presynaptic to PVN^{AVP} (**e**; female: 15 out of 364 Gal⁺ neurons, $n = 3$; male: 46 out of 180 Gal⁺ neurons, $n = 3$), to PVN^{OXT} (**f**; female: 26 out of 71 Gal⁺ neurons, $n = 3$; male: 7 out of 51 Gal⁺ neurons, $n = 3$) and to PVN^{CRH} neurons (**g**; female: 19 out of 72 Gal⁺ neurons, $n = 3$; male: 22 out of 45 Gal⁺ neurons, $n = 3$). Significantly more MPOA neurons presynaptic to PVN^{AVP} and PVN^{CRH} neurons were Gal⁺ in males than in females ($P < 0.0001$ and $P = 0.0170$, respectively, two-tailed Fisher's exact test), whereas more MPOA neurons presynaptic to PVN^{OXT} neurons

were Gal⁺ in females than in males ($P = 0.0068$). **h**, Labelling strategy for MPOA^{Gal} projections; example of retrogradely labelled Fos⁺ neuron in the MPOA. **i**, Activated fraction of MPOA^{Gal} neurons projecting to parenting-relevant brain areas ($n = 7, 4, 3, 4, 3, 4, 3, 4, 4, 4$ mice, from top to bottom). Data are mean \pm s.e.m. Red line, population average³. Projections chosen for further functional studies are labelled in blue. **j**, Strategy for monosynaptic retrograde tracing from projection-defined MPOA^{Gal} subpopulations. **k**, **l**, Map of monosynaptic inputs into VTA-projecting MPOA^{Gal} neurons (**k**) and matrix displaying inputs into projection-defined MPOA^{Gal} subpopulations (**l**; see Methods; $n = 5, 3, 4, 4, 4, 5, 5, 4, 4, 3$ mice, from top to bottom). A Tukey post hoc test was used to assess whether candidate projections (blue) receive quantitatively different inputs. VTA versus PAG, $*P = 0.0205$; PAG versus PVN, $***P = 0.0002$; all other comparisons, $***P < 0.0001$. Scale bars, 500 μ m (**b**, left) 250 μ m (**b**, inset) and 50 μ m (**e–g**, **h**).

can only infect neurons that project to an area of choice¹² (Fig. 2j and Extended Data Fig. 5b–d). We found that MPOA^{Gal} projections integrate broad input combinations, with characteristic sets of enriched or depleted inputs (Fig. 2k, l). This is seen for projections from the PAG, MeA, PVN and VTA, which receive similar, albeit quantitatively different, inputs (Fig. 2l). Notably, inputs from the nucleus accumbens and lateral septum, areas involved in reward and emotional responses, respectively, were specifically enriched in VTA-projecting MPOA^{Gal} neurons (Fig. 2k, l). Together, these findings suggest a circuit architecture in which broad input combinations converge onto largely non-overlapping, projection-defined MPOA^{Gal} subpopulations. These subpopulations may in turn be differentially activated during parenting by integrating across quantitatively different sets of activated inputs.

Specific activity of MPOA^{Gal} pools

We next used fibre photometry^{13,14} (Fig. 3a, b) to investigate whether individual MPOA^{Gal} subpopulations are active during specific parenting steps. Conditional expression of the calcium reporter GCaMP6m in MPOA^{Gal} neurons was achieved by viral injection (Extended Data Fig. 6a) and an optical fibre was implanted above the injection site (Extended Data Fig. 6b–d). The entire (pan-MPOA^{Gal}) population displayed high activity during all pup-directed parenting episodes in mothers, virgin females and fathers (Fig. 3c–g and Supplementary Video 1), but not during non-pup-directed (nest building) or passive

(crouching) parenting episodes (Fig. 3h, i). MPOA^{Gal} activation was stimulus-specific: interactions with adults resulted in minimal activity (Extended Data Fig. 6k, l). Moreover, orofacial motor actions similar to pup interactions did not activate MPOA^{Gal} neurons, confirming that the observed signals were not motion-related. The tuning of MPOA^{Gal} neurons during parenting was similar in all three groups (Fig. 3q)—highlighting their common role in the control of parental interactions. Activation during pup sniffing was higher in mothers than in virgin females and fathers (Fig. 3c), possibly reflecting the very high sensitivity of postpartum females to pup stimuli¹⁵ (Extended Data Fig. 7). Furthermore, activity decreased in mothers—but not in fathers—during eating, self-grooming and sniffing of food (Fig. 3j–l). MPOA^{Gal} neurons receive their second-largest fractional input from the arcuate nucleus, a feeding control centre¹⁶ (Fig. 1c and Extended Data Fig. 1a), suggesting that inhibition from circuits controlling mutually exclusive motor patterns, such as eating and pup grooming, might cause this decrease in activity.

To record the activity of projection-defined MPOA^{Gal} subpopulations, we injected MPOA^{Gal} target areas with a Cre-dependent, GCaMP6-expressing herpes simplex virus and implanted an optical fibre above the retrogradely labelled cell bodies (Fig. 3m and Extended Data Fig. 6e–h). PAG-projecting MPOA^{Gal} neurons were specifically activated during pup grooming (Fig. 3n and Extended Data Fig. 6m–q), whereas MeA-projecting MPOA^{Gal} neurons were active during most

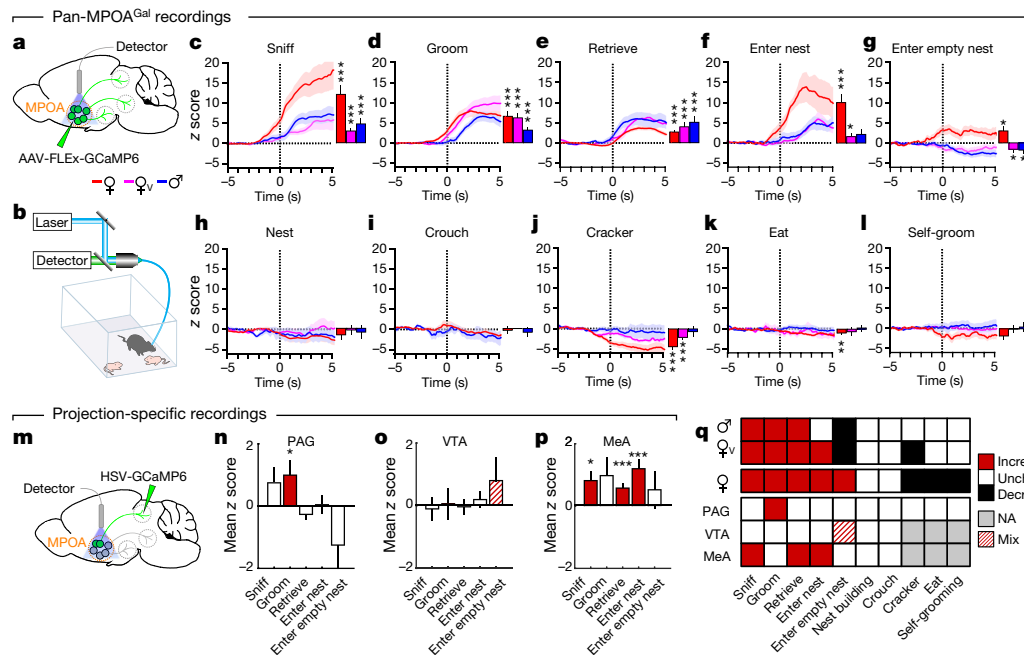


Fig. 3 | Distinct projection-defined MPOA^{Gal} neuronal pools are tuned to specific aspects of parental behaviour. **a, b**, Fibre photometry recording strategy (**a**) and setup (**b**). **c–l**, Averaged recording traces from MPOA^{Gal} population activity during pup sniffing (**c**), pup grooming (**d**), pup retrieval (**e**), entering a nest with pups (**f**), entering an empty nest (**g**), nest building (**h**) and crouching (**i**). Red, mother; pink, virgin female; blue, male. Mean peak activity (z scores) shown in mothers ($n=4$), virgin females ($n=3$) and fathers ($n=5$). **j–l**, Averaged recording traces and mean peak activity during control behaviours. Cracker indicates sniffing of a pup-sized food object. **m**, Strategy for recording projection-defined MPOA^{Gal} subpopulations. **n–p**, Mean peak activation for MPOA^{Gal}

neurons projecting to PAG (**n**, $n=10$ mice), VTA (**o**, $n=12$ mice) and MeA (**p**, $n=8$ mice) during parenting. **q**, Tuning matrix for pan-MPOA^{Gal} (top) and projection-specific (bottom) recordings. Red, increased; white, unchanged; black, decreased; NA, not available (grey). Two-tailed *t*-tests (Methods). **c**, *** $P<0.0001$, *** $P<0.0001$, *** $P=0.0001$ (from left to right); **d**, *** $P<0.0001$; **e**, *** $P<0.0001$, *** $P=0.0008$, *** $P=0.0004$ (from left to right); **f**, *** $P<0.0001$, * $P=0.0247$; **g**, * $P=0.0185$, * $P=0.0365$, * $P=0.0105$ (from left to right); **j**, *** $P=0.0002$, *** $P<0.0001$ (from left to right); **k**, ** $P=0.0059$; **n**, * $P=0.0362$; **p**, * $P=0.0102$, *** $P<0.0001$, *** $P=0.0001$ (from left to right). Data are mean \pm s.e.m.

episodes of parental behaviour (Fig. 3p and Extended Data Fig. 6m–q), indicating a more general role in parenting. Consistent with their weak Fos activation after parenting (Fig. 2i), no significant activity changes were detected in VTA-projecting MPOA^{Gal} neurons (Fig. 3o and Extended Data Fig. 6m–p). Nevertheless, MPOA^{Gal} neurons signalling to VTA neurons were weakly responsive during nest entering in a subset of animals (Fig. 3o and Extended Data Fig. 6q; 4 out of 12 mice), potentially reflecting the expectation or drive to interact with pups. Taken together, these findings support the idea that MPOA^{Gal} neurons form functionally distinct modules that are tuned to specific parenting episodes.

Functionally distinct MPOA^{Gal} pools

We tested the hypothesis that MPOA^{Gal} neurons form functionally specialized pools by optogenetically activating projections to PAG, VTA and MeA during pup interactions (Fig. 4a). We virally expressed channelrhodopsin-2 (ChR2) in MPOA^{Gal} neurons (Extended Data Fig. 8a), and implanted optical fibres above MPOA^{Gal} projection targets. Optogenetic activation of MPOA^{Gal} to PAG projections at axon terminals did not affect the fraction of parental virgin females but suppressed pup attacks in infanticidal virgin males (Fig. 4b), and—consistent with MPOA^{Gal} to PAG activity during parenting (Fig. 3n)—increased pup grooming and pup-directed sniffing bouts in both males and females (Fig. 4c and Extended Data Fig. 8c). Next, we assessed the motivation to interact with pups by inserting a climbable barrier in the home cage between the test animal and pups (Fig. 4d). Activation of MPOA^{Gal} to PAG projections had no effect on the number of barrier crosses (Fig. 4d). Importantly, the effects of activation of MPOA^{Gal} to PAG projections were specific to pup interactions, and did not affect interactions with adult conspecifics (Fig. 4e, f).

By contrast, activation of MPOA^{Gal} to VTA projections did not affect pup interactions (Fig. 4g, h), but increased barrier crossing in both

males and females (Fig. 4i and Supplementary Video 2), indicating an increased motivation to interact with pups. Interestingly, virgin males still exhibited pup-directed aggression after crossing the barrier, suggesting that this effect is not contingent upon the display of parenting. Nevertheless, in naturalistic situations, MPOA^{Gal} neurons and associated VTA projections are activated exclusively during parental interactions, thus specifically mediating parental drive. MPOA^{Gal} to VTA activation did not increase locomotion (Extended Data Fig. 8j, k) and did not affect interactions with intruders of either sex (Fig. 4j, k).

Finally, activation of MPOA^{Gal} to MeA projections did not affect pup-directed behaviours (Fig. 4l, m and Extended Data Fig. 7f, g)—except for a decrease in the amount of time spent in the nest in the females (Extended Data Fig. 8f)—or the motivation to interact with pups (Fig. 4n). However, this manipulation significantly inhibited male–male aggression and chemoinvestigation of a male intruder in females (Fig. 4o, p). Thus, instead of directly influencing parental behaviour, MPOA^{Gal} to MeA activation inhibits social interactions with adult conspecifics.

We tested the necessity of these subpopulations for discrete behaviours by expressing the inhibitory opsin eNpHR3.0 in MPOA^{Gal} neurons and stimulating their projections in virgin females (Fig. 4q, t, w). Consistent with ChR2 data, optogenetic inhibition of MPOA^{Gal} to PAG projections significantly reduced pup grooming and pup-directed sniffing bouts (Fig. 4s and Extended Data Fig. 8n), without affecting other behaviours (Fig. 4r and Extended Data Fig. 8n–p, u). By contrast, inhibition of MPOA^{Gal} to VTA projections specifically reduced barrier crossing frequency (Fig. 4v, u and Extended Data Fig. 8q, r, v), except for a reduction in time spent in the nest (Extended Data Fig. 8q). Finally, inhibition of MPOA^{Gal} to MeA projections did not affect interactions with an intruder (Fig. 4y) or other behaviours (Fig. 4x and Extended Data Fig. 8s, t, w). Recent findings indicate that representations of social stimuli in MeA and hypothalamic centres

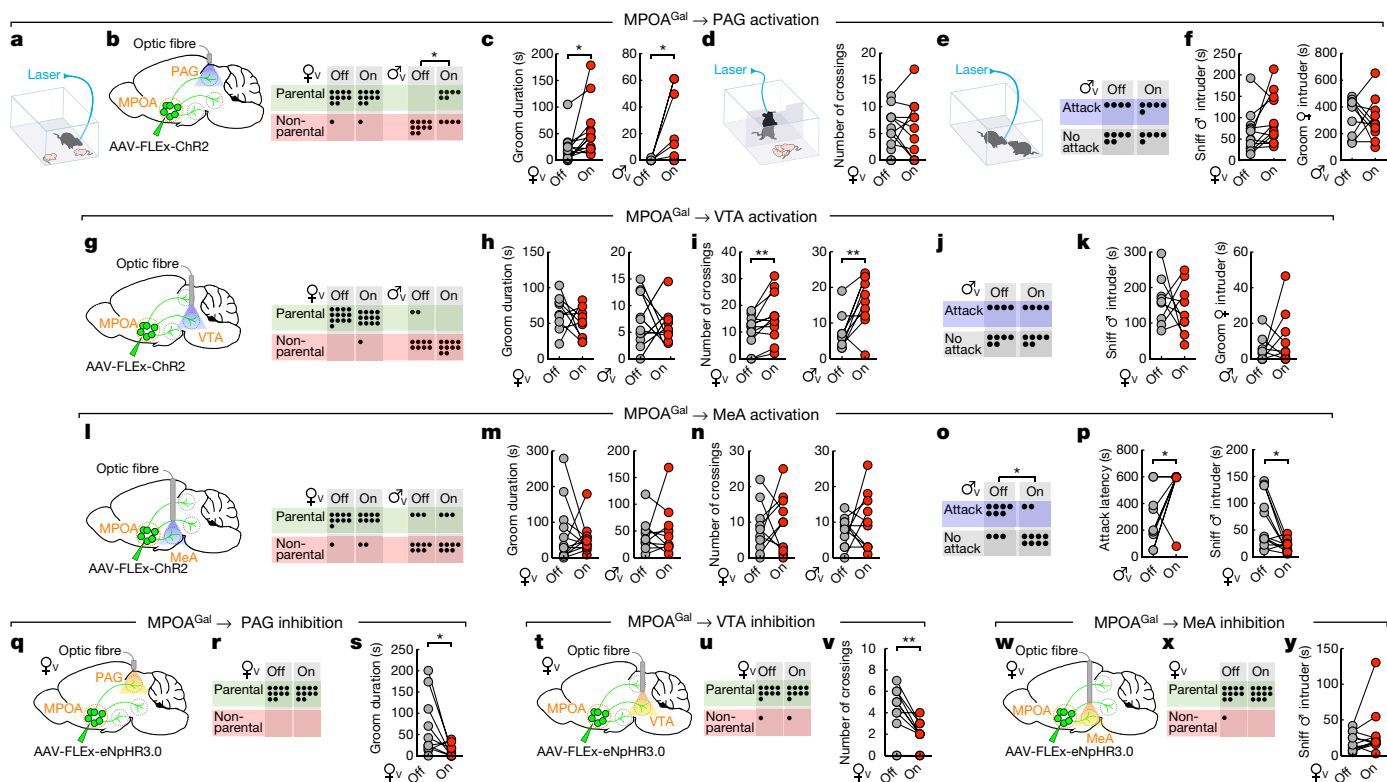


Fig. 4 | MPOA^{Gal} projections mediate discrete aspects of parental behaviour. **a**, Setup for optogenetic manipulations. **b**, **g**, **l**, Left, activation of MPOA^{Gal} projections. Right, pup-directed behaviour in virgin females and males without (Off) or with (On) activation of MPOA^{Gal} to PAG (**b**), VTA (**g**) and MeA (**l**) projections. Dots indicate the number of animals. **c**, **h**, **m**, Effect of activating MPOA^{Gal} to PAG (**c**; $n = 13$ virgin females; $n = 9$ virgin males), VTA (**h**; $n = 9$ virgin females; $n = 10$ virgin males) or MeA (**m**; $n = 10$ virgin females; $n = 10$ virgin males) projections on pup grooming. **d**, Motivation assay. **e**, **i**, **n**, Effect of activating MPOA^{Gal} to PAG (**d**; $n = 13$ virgin females), VTA (**i**; $n = 10$ virgin females; $n = 13$ virgin males) or MeA (**n**; $n = 10$ virgin females; $n = 10$ virgin males) projections on barrier crossing. **e**, Intruder assay. **e**, **j**, **o**, Effect of activating MPOA^{Gal} to PAG (**e**; $n = 10$ virgin males), VTA (**j**; $n = 10$ virgin males) or MeA (**o**; $n = 10$ virgin males) projections on male-male aggression. **f**, **k**, Effect of activating PAG (**f**) or VTA (**k**) projections on male- ($n = 12$ virgin females (**f**),

$n = 9$ virgin females (**k**) or female-directed ($n = 10$ virgin males (**f**), $n = 10$ virgin males (**k**) behaviour. **p**, Effect of activating MPOA^{Gal} to MeA projections on male-directed attack latency ($n = 10$ virgin males) and chemoinvestigation ($n = 10$ virgin females). **q**, **t**, **w**, Inhibition of MPOA^{Gal} projections. **r**, **u**, **x**, Pup-directed behaviour in virgin females without (Off) or with (On) inhibition of PAG (**r**; $n = 10$), VTA (**u**; $n = 10$) and MeA (**x**; $n = 11$) projections. **s**, Effect of inhibiting MPOA^{Gal} to PAG projections on pup grooming ($n = 10$). **v**, Effect of inhibiting MPOA^{Gal} to VTA projections on barrier crossing ($n = 10$). **y**, Effect of inhibiting MPOA^{Gal} to MeA projections on male-directed chemoinvestigation ($n = 11$). χ^2 tests (**b**, **e**, **g**, **j**, **l**, **o**, **r**, **u**, **x**) or two-tailed paired t -tests (**c**, **d**, **f**, **h**, **i**, **k**, **m**, **n**, **p**, **s**, **v**, **y**) were used. **b**, $**P = 0.0034$; **c**, $*P = 0.0273$, $*P = 0.0374$; **i**, $**P = 0.0089$, $**P = 0.0056$; **o**, $*P = 0.0246$; **p**, $*P = 0.033$, $*P = 0.0109$; **s**, $*P = 0.0396$; **v**, $**P = 0.0038$.

change significantly after sexual experience^{17,18}. Thus, low basal activity in this circuit branch in virgin females compared to mothers may preclude further inhibition. Alternatively, or additionally, this lack of effect may result from a more complex role of the connectivity from MPOA^{Gal} neurons projecting to MeA.

Concluding remarks

Taken together, our data suggest that distinct MPOA^{Gal} pools control discrete aspects of parental behaviour in both sexes (Fig. 5). Consistent with a role of the PAG in motor aspects of maternal behaviour², MPOA^{Gal} to PAG projections promote pup grooming. Retrograde tracing from PAG showed that MPOA^{Gal} neurons synapse with GABAergic (γ -aminobutyric-acid-releasing, inhibitory), but not glutamatergic (excitatory) PAG neurons (Extended Data Fig. 2h–j). Because the vast majority (around 90%) of MPOA^{Gal} neurons are GABAergic³, pup grooming is probably elicited by disinhibition in the PAG. Indeed, infusion of the PAG with the GABA_A receptor antagonist bicuculline increases pup licking and grooming¹⁹. By contrast, MPOA^{Gal} to VTA projections specifically influence the motivation to interact with pups without affecting the quality of adult–infant interactions. This is consistent with the proposed role of the VTA in motivation²⁰ and social reinforcement²¹, and complements previous findings in rats^{2,22}. Nearby Gal⁺ neurons in the lateral hypothalamus promote food-seeking behaviour, despite lacking VTA projections²³, further highlighting the

specific role of MPOA^{Gal} neurons in parenting. Finally, we found that MPOA^{Gal} to MeA projections do not directly influence pup-directed behaviour, but instead inhibit potentially competing adult social interactions.

Interestingly, MPOA^{Gal} to MeA projections are active during most episodes of parenting (Fig. 3p, q), suggesting that the entire behaviour, rather than specific parenting components, are broadcast by this projection to influence the vomeronasal pathway^{24–26}. Specific inhibitory feedback from MPOA^{Gal} to MeA projections might impair the detection, or alter the valence, of non-pup-related social stimuli. Indeed, optogenetic stimulation of glutamatergic neurons in the posteriodorsal MeA—the MeA compartment that is most densely innervated by MPOA^{Gal} fibres (Fig. 2b)—has been shown to suppress interactions with adult conspecifics²⁷. The projections investigated here mediate crucial, non-overlapping aspects of parental behaviour and the sum of their activity profiles matches that of the entire MPOA^{Gal} population (Fig. 3q). Thus, combined with the finding that MPOA^{Gal} neurons contact AVP-, OXT- and CRH-expressing PVN neurons (Fig. 2e–g), we have dissected circuit branches for four major—motor, motivational, social and neuromodulatory—aspects of parenting control. Other MPOA^{Gal} projections that have not been included here may have additional roles in parenting. Lastly, our tracing data suggest extensive connectivity within the MPOA (Fig. 1c), hinting at interactions between functionally specialized MPOA^{Gal} subpopulations.

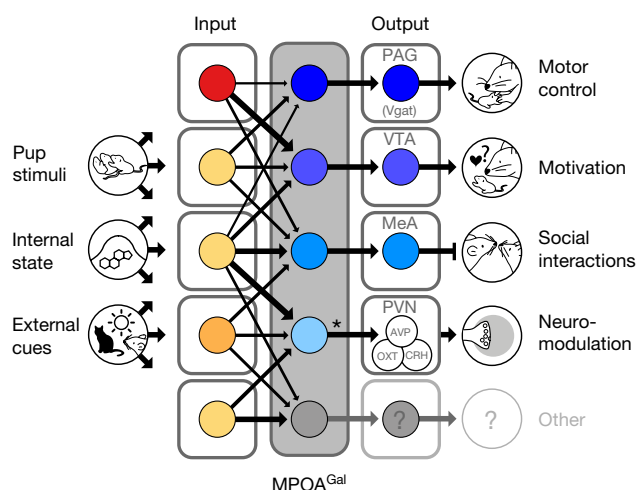


Fig. 5 | Functional architecture of the MPOA^{Gal} circuit. Broad, state- and sex-specifically activated inputs converge onto largely non-overlapping, projection-defined MPOA^{Gal} subpopulations that elicit specific aspects of parental behaviour. *MPOA^{Gal} to PVN connections are sexually dimorphic (see Fig. 2e–g).

Considerable progress has recently been made in identifying neuronal populations that control specific social behaviours or homeostatic functions^{10,16,28–31}. However, little is known about how these multi-component behaviours or functions are orchestrated at the circuit level. Intriguingly, the modular architecture uncovered here for the control of parenting is reminiscent of the motor circuit motif that has been identified in the mammalian spinal cord, in which discrete phases of locomotor sequences are controlled by functionally distinct neuronal pools with highly specific connectivity patterns³². Whether other social behaviours rely on similar circuit architectures remains to be determined.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0027-0>.

Received: 4 August 2017; Accepted: 13 February 2018;

Published online 11 April 2018.

- Dulac, C., O'Connell, L. A. & Wu, Z. Neural control of maternal and paternal behaviors. *Science* **345**, 765–770 (2014).
- Numan, M. & Insel, T. R. *The Neurobiology of Parental Behavior* (Springer, New York, 2011).
- Wu, Z., Autry, A. E., Bergan, J. F., Watabe-Uchida, M. & Dulac, C. G. Galanin neurons in the medial preoptic area govern parental behaviour. *Nature* **509**, 325–330 (2014).
- Wickersham, I. R. et al. Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
- Marlin, B. J., Mitre, M., D'Amour, J. A., Chao, M. V. & Froemke, R. C. Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* **520**, 499–504 (2015).
- Johnson, Z. V. & Young, L. J. Oxytocin and vasopressin neural networks: Implications for social behavioral diversity and translational neuroscience. *Neurosci. Biobehav. Rev.* **76**, 87–98 (2017).
- Bendesky, A. et al. The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434–439 (2017).
- Tachikawa, K. S., Yoshihara, Y. & Kuroda, K. O. Behavioral transition from attack to parenting in male mice: a crucial role of the vomeronasal system. *J. Neurosci.* **33**, 5120–5126 (2013).
- Vom Saal, F. S. Time-contingent change in infanticide and parental behavior induced by ejaculation in male mice. *Physiol. Behav.* **34**, 7–15 (1985).
- Kohl, J., Autry, A. E. & Dulac, C. The neurobiology of parenting: a neural circuit perspective. *BioEssays* **39**, 1–11 (2017).
- Scott, N., Prigge, M., Yizhar, O. & Kimchi, T. A sexually dimorphic hypothalamic circuit controls maternal care and oxytocin secretion. *Nature* **525**, 519–522 (2015).
- Schwarz, L. A. et al. Viral-genetic tracing of the input–output organization of a central noradrenaline circuit. *Nature* **524**, 88–92 (2015).

- Gunaydin, L. A. et al. Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535–1551 (2014).
- Kudo, Y. et al. A single optical fiber fluorometric device for measurement of intracellular Ca²⁺ concentration: its application to hippocampal neurons *in vitro* and *in vivo*. *Neuroscience* **50**, 619–625 (1992).
- Elyada, Y. M. & Mizrahi, A. Becoming a mother—circuit plasticity underlying maternal behavior. *Curr. Opin. Neurobiol.* **35**, 49–56 (2015).
- Andermann, M. L. & Lowell, B. B. Toward a wiring diagram understanding of appetite control. *Neuron* **95**, 757–778 (2017).
- Li, Y. et al. Neuronal representation of social information in the medial amygdala of awake behaving mice. *Cell* **171**, 1176–1190 (2017).
- Remedios, R. et al. Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. *Nature* **550**, 388–392 (2017).
- Lee, G. & Gammie, S. C. GABA_A receptor signaling in caudal periaqueductal gray regulates maternal aggression and maternal care in mice. *Behav. Brain Res.* **213**, 230–237 (2010).
- Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).
- McHenry, J. A. et al. Hormonal gain control of a medial preoptic area social reward circuit. *Nat. Neurosci.* **20**, 449–458 (2017).
- Seip, K. M. & Morrell, J. I. Transient inactivation of the ventral tegmental area selectively disrupts the expression of conditioned place preference for pup- but not cocaine-paired contexts. *Behav. Neurosci.* **123**, 1325–1338 (2009).
- Qualls-Creekmore, E. et al. Galanin-expressing GABA neurons in the lateral hypothalamus modulate food reward and noncompulsive locomotion. *J. Neurosci.* **37**, 6053–6065 (2017).
- Isogai, Y. et al. Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241–245 (2011).
- Bergan, J. F., Ben-Shaul, Y. & Dulac, C. Sex-specific processing of social cues in the medial amygdala. *eLife* **3**, e02743 (2014).
- Yao, S., Bergan, J., Lanjuin, A. & Dulac, C. Oxytocin signaling in the medial amygdala is required for sex discrimination of social cues. *eLife* **6**, e31373 (2017).
- Hong, W., Kim, D. W. & Anderson, D. J. Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala neuronal subsets. *Cell* **158**, 1348–1361 (2014).
- Anderson, D. J. Circuit modules linking internal states and social behaviour in flies and mice. *Nat. Rev. Neurosci.* **17**, 692–704 (2016).
- Yang, T. & Shah, N. M. Molecular and neural control of sexually dimorphic social behaviors. *Curr. Opin. Neurobiol.* **38**, 89–95 (2016).
- Zimmerman, C. A., Leib, D. E. & Knight, Z. A. Neural circuits underlying thirst and fluid homeostasis. *Nat. Rev. Neurosci.* **18**, 459–469 (2017).
- Weber, F. & Dan, Y. Circuit-based interrogation of sleep control. *Nature* **538**, 51–59 (2016).
- Arber, S. Motor circuits in action: specification, connectivity, and function. *Neuron* **74**, 975–989 (2012).

Acknowledgements We thank S. Sullivan for help with behaviour and mouse husbandry. E. Kremer (Montpellier) and R. Neve (MIT) provided viral vectors. E. Soucy and J. Greenwood helped design motivation assay. R. Hellmiss and K. Wilbur helped with illustrations. H. S. Knobloch-Bollmann provided advice on PVN cell types. We thank the members of the Dulac laboratory for comments on the manuscript. This work was supported by a Human Frontier Long-Term Fellowship, an EMBO Long-Term Fellowship and a Sir Henry Wellcome Fellowship to J.K., Fondation pour la Recherche Médicale grant SPE20150331860 to B.M.B., a NIH K99 Award and a NARSAD Young Investigator Award to A.E.A., a Howard Hughes Gilliam Fellowship to B.M.-R., a Harvard Mind Brain and Behavior faculty grant to N.U. and NIH grant 1R01HD082131-01A1 to C.D. C.D. and L.L. are investigators of the Howard Hughes Medical Institute.

Reviewer information Nature thanks V. Grinevich and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.K. and C.D. conceived and designed the study. J.K. performed and analysed tracing and optogenetics experiments. J.K. and B.M.B. performed and analysed photometry recordings. A.E.A. helped with in situ hybridization experiments. B.M.-R. helped with CTB tracing experiments. N.D.R. helped to analyse tracing data, and V.K. helped with optogenetics experiments. L.S.Z., K.M. and L.L. shared unpublished viral reagents. N.U. provided fibre photometry setup. J.K., B.M.B., N.D.R., A.E.A. and C.D. analysed and interpreted the results. J.K. and C.D. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0027-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0027-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to C.D. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Animals. The *Gal::cre* BAC transgenic line (Stock: Tg(Gal-cre)KI87Gsat/Mmucd, 031060-UCD) was imported from the Mutant Mouse Regional Resource Center and has previously been described³. Cre-dependent tdTomato reporter mice (*Gt(Rosa)26So^{tm9}(CAGtdTomato)Hze*)³³, C57BL/6J, OXT-IRES-Cre, Vgat-IRES-Cre and TH-IRES-Cre mice were obtained from Jackson Laboratories. Vglut2-IRES-Cre mice were provided by B. Lowell. The AVP-IRES-Cre line has previously been described⁷. CRH-IRES-Cre mice were obtained from B. Lowell, J. Majzoub and Jackson Laboratories. Animals were maintained on 12h:12h light:dark cycle (light on: 02:00–14:00) with food and water available ad libitum. Animal care and experiments were carried out in accordance with the NIH guidelines and approved by the Harvard University Institutional Animal Care and Use Committee (IACUC).

Histology and immunostaining. Animals were perfused transcardially with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA) in PBS. Brains were dissected and post-fixed in 4% PFA for 16h, then washed in PBS for 6h. After embedding in 4% low-melting point agarose (Thermo Fisher, 16520-050) in PBS, 60- μ m coronal sections were cut on a vibratome (Leica) and mounted on Superfrost Plus slides (VWR, 48311-703) with DAPI-containing VECTASHIELD mounting medium (Vector Laboratories, H-1200). For immunostaining in 48-well culture plates, sections were permeabilized for 30 min in PBS-T (0.3% Triton X-100 in PBS), post-fixed with PFA for 10 min, and washed in PBS-T (three times, 20 min each). Blocking was carried out overnight in blocking buffer (0.3% Triton X-100, 1% BSA, 2% normal donkey serum in PBS). Incubation with primary antibodies was performed for 24–48 h on a Nutator at 4 °C. After washing in PBS-T (five times, 60 min each), secondary antibodies were added for 48 h at 4 °C. After final washes in PBS-T (five times, 60 min each), sections were mounted. Primary antibodies: goat anti-Fos (Santa Cruz, sc-52, 1:500), chicken anti-GFP (Abcam, ab13970, 1:1,000), rabbit anti-AVP (Immunostar, 20069, 1:6,000), rabbit anti-OXT (Immunostar, 20068, 1:6,000). Secondary antibodies (all from Thermo Fisher): Alexa Fluor-568 anti-goat (A-11057, 1:1,500), Alexa Fluor-555 anti-goat (A-21432, 1:1,500) and Alexa Fluor-647 anti-goat (A-21447, 1:1,500). All antibodies were incubated in PBS-T, with the exception of Fos antibody, which was incubated in PBS.

RNA in situ hybridization. Freshly dissected brains were embedded in OCT (Tissue-Tek, 4583) and frozen with dry ice. Subsequently, 16- μ m cryosections were collected on Superfrost Plus slides (VWR, 48311-703) and used for mRNA in situ hybridization. Fluorescent mRNA in situ hybridization was performed mostly as described²⁴. Complementary DNA (cDNA) of *Gal* or *eYFP* mRNA was cloned in approximately 800-base-pair segments into a pCRII-TOPO vector (Thermo Fisher, K465040). Antisense complementary RNA (cRNA) probes were synthesized with T7 (Promega, P2075) or Sp6 polymerases (Promega, P1085) and labelled with digoxigenin (DIG, Roche 11175025910) or fluorescein (FITC, Roche 11685619910). Hybridization was performed with 0.5–1.0 ng ml⁻¹ cRNA probes at 68 °C. Probes were detected using horseradish peroxidase (POD)-conjugated antibodies (anti-FITC-POD, Roche 11426346910, 1:250; anti-DIG-POD, Roche 11207733910, 1:500). Signals were amplified using biotin-conjugated tyramide (Perkin Elmer NEL749A001KT) and subsequently visualized with Alexa Fluor-488-conjugated streptavidin (Thermo Fisher, S11223) or the TSA-plus Cy3 system (Perkin Elmer, NEL744001KT).

Viruses. Recombinant AAV vectors were produced by the UNC Vector Core. AAV titres ranged from 1.3 to 2.6 $\times 10^{12}$ viral particles ml⁻¹, based on quantitative PCR analysis. Pseudotyped, G-deleted rabies virus⁴ was obtained from the Salk vector core at a titre of 4.3 $\times 10^8$ viral particles ml⁻¹. The pAAV-CAG-FLEX-Syn-GFP plasmid was provided by S. Arber and AAV1/CAG-FLEX-Syn-GFP was produced by the UNC Vector Core. The pAAV-CAG-FLEX-TCB, pAAV-CAG-FLEX-RG³⁴, pAAV-CAG-FLEX^{FRT}-TC and pAAV-CAG-FLEX^{FRT}-RG plasmids were provided by L.L. (Stanford University), and AAV5/DJ-hSyn1-FLEX^{FRT}-mGFP³⁵, AAV1/CAG-FLEX^{FRT}-TC and AAV1/CAG-FLEX^{FRT}-RG were packaged by the UNC Vector core. L.L. and E. Kremer provided CAV2-FLEX^{loxP}-Flp. L.S.Z. provided CAV2-FLEX-ZsGreen. AAV1/CAG-FLEX-tdTomato, AAV1/Syn-FLEX-GCaMP6m, AAV5/EF1 α -DIO-hChR2(H134R)-eYFP and AAV5/EF1 α -DIO-eYFP were purchased from UPenn Vector core. HSV-hEF1 α -LSL1-GCaMP6m (HT) was obtained from MIT Vector Core.

Anterograde tracing. Anterograde tracing experiments were performed in *Gal::cre* mice (or in C57BL/6J for control experiments) at around 8–12 weeks of age. All surgeries were performed under aseptic conditions in animals anaesthetized with 100 mg kg⁻¹ ketamine (KetaVed, Vedco) and 10 mg kg⁻¹ xylazine (AnaSed) via intraperitoneal (i.p.) injection. Using a Nanoject II injector (Drummond Scientific), 300 nl of a 1:1 mixture of AAV1/CAG-FLEX-tdTomato:AAV1/CAG-FLEX-Syn-GFP³⁶ (synaptophysin-GFP) was injected into the MPOA (coordinates: anteroposterior (AP): 0.0 mm from Bregma; mediolateral (ML): -0.5 mm from the midline, dorsoventral (DV): -5.05 mm) to visualize presynaptic terminals of MPOA^{Gal} neurons. Syn-GFP was chosen to distinguish presynaptic sites from fibres of passage. Analgesia (buprenorphine, 0.1 mg kg⁻¹, i.p.) was administered for two

days after each surgery. Two weeks later, mice were euthanized and dissected. In some experiments, a 1:1 mixture of AAV1/CAG-FLEX-tdTomato:AAV1/CAG-FLEX-Syn-GFP was injected to visualize presynaptic terminals of MPOA^{Gal} neurons. For quantification of synaptic density, the average pixel intensity in a target region containing presynaptic GFP⁺ punctae was calculated and the background was subtracted. Because injections were unilateral and no labelling was observed in most cases contralaterally, the equivalent region on the contralateral hemisphere was chosen for background subtraction; in cases where contralateral GFP⁺ punctae were present, an adjacent unlabelled region was chosen. Background-corrected intensities were normalized to the average pixel intensity at the MPOA injection site for each brain.

Trans-synaptic retrograde tracing. Input tracing experiments were performed in *Gal::cre* mice (or C57BL/6J in control experiments) at about 8–12 weeks of age. We injected 150–200 nl of a 1:1 mixture of AAV1/CAG-FLEX-TCB:AAV1/CAG-FLEX-RG unilaterally into the MPOA. Two weeks later, 450–600 nl EnvA-pseudotyped, RG-deleted, GFP-expressing rabies virus (EnvA- Δ G-rabies) was injected into the MPOA. After recovery, mice were housed in a biosafety-level-2 (BL2) facility for four days before euthanization. Relative input strength was quantified from brain sections as follows: every second 60- μ m section was imaged and cells were counted using the ImageJ CellCounter plugin. GFP⁺ cells on the injected hemisphere were counted and assigned to brain areas based on classifications of the Paxinos Mouse Brain Atlas³⁷, using anatomical landmarks in the sections visualized by DAPI staining and tissue autofluorescence. In addition, all contralateral and non-assigned GFP⁺ cells were counted to obtain the total number of GFP⁺ cells. We then quantified the number of ipsilateral mCherry⁺ starter neurons per brain area and the total number of starter neurons. Because starter neurons are both GFP⁺ and mCherry⁺, whereas presynaptic neurons are only GFP⁺, the total number of starter neurons was subtracted from the total number of GFP⁺ neurons to obtain the total number of presynaptic neurons within the MPOA. Finally, the relative input fraction for each area was determined by dividing the number of presynaptic neurons detected in that brain area by the total number of presynaptic neurons in a given brain. Injection of starter AAVs and EnvA- Δ G-rabies into the MPOA of C57BL/6J mice did not result in detectable background labelling (Extended Data Fig. 5a). Inputs from PAG were detected only in a subset of animals. Presynaptic AVP⁺ neurons in the PVN were identified as predominantly magnocellular based on cell body size^{38,39} and position⁴⁰. Presynaptic neurons in the MPOA (Fig. 2d–g and Extended Data Fig. 2e–j) were identified as Gal⁺ by in situ hybridization.

Lateralization effects. Retrograde and anterograde tracing experiments were performed in the right hemisphere. However, a recent study found that the oxytocin receptor is more highly expressed in the left auditory cortex of females and that OXT binding there is crucial for pup retrieval⁵. We therefore investigated potential lateralization effects by tracing MPOA^{Gal} neurons in the left hemisphere. Resulting presynaptic neuron numbers and projection patterns (Extended Data Figs. 1b, 2c) were indistinguishable from those obtained after right-hemispheric tracing, suggesting that anatomical lateralization is not a dominant feature of the subcortical circuits described here.

Projection-specific trans-synaptic retrograde tracing. For projection-specific trans-synaptic retrograde tracing (cTRIO (cell-type-specifically tracing the relationship between input and output))¹², 300–500 nl of CAV2-FLEX^{loxP}-Flp was injected into identified target areas of MPOA^{Gal} neurons (for coordinates, see Extended Data Table 1) in 8–12-week-old *Gal::cre* mice. During the same surgery, 300–600 nl of a 1:1 mixture of AAV1/CAG-FLEX^{FRT}-TC:AAV1/CAG-FLEX^{FRT}-RG¹² (starter AAVs) was injected into the MPOA. This combination of Cre-dependent, Flp-expressing CAV and Flp-dependent starter AAVs renders MPOA^{Gal} neurons projecting to a specific target area susceptible to subsequent infection with G-deleted, EnvA-pseudotyped rabies virus. Two weeks later, 450–500 nl of EnvA- Δ G-rabies was injected into the same MPOA coordinate. After recovery, mice were housed in a biosafety-level-2 (BL2) facility for four days before euthanization. Injection of starter AAVs without CAV did not result in expression (Extended Data Fig. 5b, c). However, because the injection of all cTRIO tracing viruses into C57BL/6J mice resulted in background expression near the injection site (Extended Data Fig. 5d), the following areas were excluded from analysis: MPOA, BNST, AH, PVN and supraoptic nucleus (SON). This background labelling is probably due to low levels of Cre- or Flp-independent expression of TVA-mCherry and RG¹².

We quantified the connectivity of each MPOA^{Gal} projection to its inputs using a multinomial regression model (response: neuron counts in each input area, factors: MPOA^{Gal} projections). The baseline category in the model was represented by the mean input fraction across all experiments. Reported effects are therefore relative to a randomly chosen projection and the *P* values reported in Fig. 2k, l are obtained from a normal distribution in which the *z* scores are the effects of the multinomial regression divided by their corresponding standard errors. To test for differences in the multinomial distribution of input to target region projections, the least-square

means from the multinomial regression model was computed using the *lsmeans* package in R and used to run all pairwise comparisons.

MPOA^{Gal} input activity screen. To determine which fraction of MPOA^{Gal} inputs is activated during parental behaviour, viral injections were performed as described in ‘Trans-synaptic retrograde tracing’. Animals were single-housed until behavioural testing four days later with two pups (see ‘Parental behaviour assay’). For the equivalent experiments in mothers and fathers, 8–12-week-old *Gal::cre* males and females were paired up 10 days before injection of starter viruses and returned to their home cage where they remained until three days after injection of EnvA- Δ G-rabies when either the father and litter (for testing of mothers) or the mother and litter (for testing of fathers) were removed from the home cage. Parents underwent behavioural testing on the following day, that is, four days after injection of EnvA- Δ G-rabies. Typically around 80% of virgin females and more than 90% of mothers and fathers were parental. Ninety minutes after onset of retrieval, mice were deeply anaesthetized with isoflurane and rapidly perfused transcardially with 30 ml of ice-cold PBS, followed by 30 ml of ice-cold PFA (4% in PBS). Brains were dissected and post-fixed in PFA (4% in PBS) at 4 °C for 16 h. On the next day, brains were rinsed with cold PBS and 60- μ m coronal sections were prepared with a vibratome (Leica VT1000 S). Sections were further post-fixed in PFA (4% in PBS) at room temperature for 10 min and immunostaining against Fos was performed (see ‘Histology and immunostaining’). Only brains from mice that performed all steps of pup-directed parental behaviour (sniffing, retrieval, grooming, licking, crouching) were processed. Animals that were habituated in the test arena but not exposed to pups served as negative controls. Unpaired *t*-tests were used to assess activation of input areas between parental and control animals and *P* values were adjusted for multiple comparisons using the Benjamini–Hochberg method (false-discovery rate (FDR) < 0.05).

Previous studies have reported that the basic properties of Δ G-rabies-infected neurons are not altered until seven days after infection^{41,42} and likewise, effects of rabies on (transgene) expression levels have only been reported seven days after infection⁴³. Because animals were tested and perfused four days after rabies infection in our study, neuronal physiology and Fos activation should be mostly unaffected. Because we reliably observed Fos immunostaining in rabies⁺ neurons (Fig. 1g–j), rabies infection per se does not preclude activity-dependent Fos expression after four days. However, rabies infection could theoretically upregulate Fos expression in infected neurons, resulting in an overestimation of activated input neurons in our dataset. To address this possibility, we compared Fos⁺ cell numbers in the MPOA of unilaterally rabies-injected mothers between the injected (ipsilateral) and the non-injected (contralateral) hemisphere (Extended Data Fig. 1c, top). We found that numbers of Fos⁺ neurons were not significantly different between hemispheres (Extended Data Fig. 1c, bottom; *P* = 0.43; paired *t*-test; *n* = 6). Therefore, rabies infection is unlikely to strongly affect Fos⁺ expression in our experimental paradigm.

MPOA^{Gal} projection activity screen. To determine the activation of individual MPOA^{Gal} projections during parental behaviour, 300–500 nl of CAV2-FLEX-ZsGreen was injected into identified MPOA^{Gal} target areas in 8–12-week-old *Gal::cre* females. Animals were single-housed one week after injection. Behavioural testing with two pups (see ‘Parental behaviour assay’) was performed three weeks after injection to allow for efficient retrograde transport of the virus. For the equivalent experiments in fathers, 8–10-week-old *Gal::cre* virgin males were individually paired up with females for four days, injected and subsequently returned to the female. Two to three days after pups were born (around three weeks after injection), and one day before testing, the female and pups were removed from the cage. Testing, brain collection and immunostaining were performed as described in ‘MPOA^{Gal} input activity screen’. Because MPOA^{Gal} neurons are not activated in non-pup-exposed mice³, negative controls were not performed in these experiments.

Axon collateralization experiments. In order to assess axon collateralization of MPOA^{Gal} neurons (Extended Data Fig. 4), *Gal::cre* mice received injections of 300–500 nl of CAV-FLEX^{loxP}-Flp into an MPOA^{Gal} target site (for coordinates, see Extended Data Table 1) and of 600 nl of AAV5/hSyn1-FLEX^{FRT}-mGFP into the MPOA. Mice were euthanized eight weeks later and the signal was amplified by anti-GFP immunostaining.

CTB tracing. Mice expressing tdTomato in Gal⁺ neurons (*Gal::cre^{+/+}-loxP-Stop-loxP-tdTomato^{+/+}*) received pairwise injections of 50–100 nl of 0.5% (wt/vol) fluorescently labelled cholera toxin B subunit (CTB-488, Thermo Fisher C22841, CTB-647, Thermo Fisher C34778). After seven days, brains were collected, fixed and 60- μ m sections prepared. Individual sections were fixed again in 4% PFA for 10 min. The fraction of double-labelled, tdTomato⁺, Gal⁺ neurons in the MPOA was quantified. In control experiments, a 1:1 mixture of CTB-488 and CTB-647 was injected into MeA or PAG.

Imaging and image analysis. Samples were imaged using an Axio Scan.Z1 slide scanner (Zeiss), and confocal stacks were acquired on an LSM 880 confocal microscope (Zeiss). Image processing was performed using custom routines for the Fiji

distribution of ImageJ. For most tracing experiments, every second section was imaged, with the exception of MPOA^{Gal} projection activity and CTB-tracing experiments, where every MPOA-containing section was imaged and analysed.

Parental behaviour assay. Before behavioural testing animals were housed individually for 5–7 days unless otherwise specified. Experiments started at the beginning of the dark phase and were performed under dim red light. Testing was performed in the home cage (with the exception of locomotion assays, see below) and preceded by a 30-min habituation period. Two 1–4-day-old C57BL/6J pups were placed in different corners opposite the nest. Once retrieval occurred, a timer was started. Each test was recorded using a multi-camera surveillance system (GeoVision GV-1480) and behaviours were scored by an individual blind to the genotype using the Observer 5.0 or XT 8 software (Noldus Information Technology).

Fibre photometry. Fibre photometry (fluorometry) was performed as previously described⁴⁴. For photometry recordings, 8–12-week-old *Gal::cre^{+/+}-loxP-Stop-loxP-tdTomato^{+/+}* mice were used. For pan-MPOA^{Gal} recordings, 400–500 nl of AAV1/Syn-FLEX-GCaMP6m (Upenn Vector Core) was injected into the MPOA; for projection-specific recordings, 600–700 nl of hEF1 α -LS1L-GCaMP6m, a Cre-dependent, retrograde, long-term herpes simplex virus (LT-HSV) was bilaterally injected into MPOA^{Gal} target areas. During the same surgery, a custom 400- μ m fibre-optic cannula (Doric Lenses) was implanted into the MPOA (for coordinates, see Extended Data Table 1). For recordings in mothers and fathers, animals were paired up five days before surgery, to ensure that pups were born around three weeks after virus injection. One day after surgery, animals were returned to their mating partner. The implanted animal’s mating partner and offspring were removed 3–5 h before recordings. Virgin female mice were single-housed seven days before the first recording session and thereafter between experiments. Recordings were made 2–4 weeks after the surgery under IR illumination in the home cage of the mouse. Mice were briefly (around 10 min) habituated in the recording setup before 8–10 pups (1–4 days old) were introduced into the cage. Recording sessions typically lasted 10–20 min, with at least two days between sequential recordings. The implant was coupled to a custom patch cord (Doric Lenses) to simultaneously deliver 473-nm excitation light from a DPSS laser (Opto Engine LLC), passed through a neutral density filter (4.0 optical density, Thorlabs) and to collect fluorescence emission. Activity-dependent fluorescence emitted by cells in the vicinity of the implanted fibre tip was collected by a 0.65 NA microscope objective (Olympus), spectrally separated from the excitation light using a dichroic mirror (Chroma), passed through a band pass filter (ET500/50, Chroma) and focused onto a photodetector (FDS10X10, Thorlabs) connected to a current preamplifier (SR570, Stanford Research Systems). Another band pass filter (ET600/20) in front of a second photodetector/preamplifier was used to collect tdTomato fluorescence. Owing to considerable bleed-through of the GCaMP signal into the tdTomato channel, we chose not to use the tdTomato recording trace to normalize our data, instead opting for a set of behavioural controls for motion artefacts (see below). The preamplifier output voltage signal was collected by a NIDAQ board (PCI-e6321, National Instruments) connected to a computer running LabVIEW (National Instruments) for signal acquisition. Video recordings were acquired at 15 frames per second and the signal from the optical fibre was sampled at 1 kHz. A TTL-triggered photodiode next to the cage was used to align videos and voltage recording traces.

Analysis was performed using custom MATLAB (MathWorks) routines. Only recordings with a stable baseline were included in our analysis. The raw signal over each entire recording session was divided by the mean of a Gaussian fit to the distribution of GCaMP to normalize the baseline over the recording session. Since the increase in GCaMP signal preceded event detection in some cases (for example, see Fig. 3c), *z* scores were calculated using the period from –5 to –2 s before event detections as baseline and from 0 to 3 s from event detection as signal. For statistical analyses (that is, *t*-tests, ANOVA), we considered a value of *P* ≤ 0.05 significant. Behaviours were scored manually off-line by an experimenter blind to the photometry recording data. The responses to a stimulus type within a session (typically 5–10 trials per behaviour type) were averaged, and these session averages across mice were used as data displayed in Fig. 3 and Extended Data Fig. 6.

We performed a set of behavioural controls to address the possible contribution of motion artefacts to the recorded signal. In all of the following cases, (orofacial) motor actions highly identical to pup interactions did not result in detectable increases in GCaMP fluorescence intensity. No increase in signal was observed when animals retrieved or sniffed a pup-sized cracker (Fig. 3j), during eating (Fig. 3k) or during self-grooming (Fig. 3l). In addition, no increase in signal was detectable when animals retrieved bedding material to the nest (Fig. 3h). Finally, chemoinvestigation of accessible versus inaccessible pups resulted in different GCaMP responses (from –5 to 0 s period before sniffing, Extended Data Fig. 6i, j). Therefore, the increases in signal intensity observed during pup interactions very probably represent actual activity changes rather than motion artefacts.

Optogenetics. *Gal::cre* mice 8–12 weeks of age were used in these experiments. Because potential increases in parental behaviour would be difficult to detect in already highly parental mothers and fathers, we performed these experiments in virgin animals, in which a higher dynamic range of parental interactions can be assessed. Animals were exposed to two pups in their home cage (see ‘Parental behaviour assay’) and those that attacked (virgin males) or initiated parental behaviour (virgin females) within 15 min were selected for surgery. We injected 700 nl of AAV5/EF1 α -DIO-hChR2(H134R)-eYFP (activation) or AAV5/EF1 α -DIO-eNpHR3.0-eYFP (inhibition) bilaterally into the MPOA and in the same surgery a dual fibre-optic cannula (300 μ m, 0.22 NA, Doric Lenses) was implanted 0.4–0.5 mm above the respective MPOA^{Gal} projection target (Extended Data Table 1) and affixed to the skull with dental cement. Mice were tested 3–5 weeks after injection to allow for efficient expression of ChR2 or eNpHR3.0 into axon terminals. On testing day, the implant was connected to an optical fibre attached to either a 473-nm laser (150 mW, Laserglow Technologies) or a 460-nm LED (50 W, Prizmatix) for optogenetic activation, or a 589-nm laser (300 mW, Opto Engine LLC) for inhibition, via a commutator. Animals were tested in either stimulation or non-stimulation trials in randomized order, with two days between trials. In addition, the order in which animals were tested during each experimental session was randomized. In pup exposure experiments, two C57BL/6J pups, 1–3 days of age, were introduced to the test animal's home cage in each corner furthest from the nest after 10 min of habituation. For activation experiments, blue light (473 nm) was delivered in 20-ms pulses at 20 Hz for 1–4 s whenever the animal contacted a pup with its snout. The light power exiting the fibre tip was 5 mW, which we calculated as providing an irradiance of 5–10 mW mm⁻² at the target region (using the brain tissue light transmission calculator provided by the Deisseroth laboratory, <http://www.stanford.edu/group/dlab/cgi-bin/graph/chart.php>). For loss-of-function experiments, constant yellow light (589 nm) was delivered at 8–10 mW at the fibre tip, amounting to an estimated irradiance of 15–20 mW mm⁻² at the target. Each trial lasted up to 10 min but when virgin males attacked and wounded a pup, the trial was ended and the pup was euthanized.

The following behaviours were scored and quantified: pup sniffing, grooming and licking, pup retrieval to the nest, aggression (animal grabs the pup violently and attempts to bite), crouching (animal hovers above the pup in the nest), nest building and time spent in the nest. For the motivation assay, following a 10-min habituation period a transparent barrier was inserted into the home cage, dividing the cage into a nest and a pup compartment. Next, 4–5 pups were introduced into the pup compartment and 473-nm light was delivered in 20-ms pulses at 20 Hz for 4 s every 10 s for a total of 6 min. Locomotion was assessed in a 36 \times 25-cm arena over a period of 5 min. In stimulation trials, 473 nm light (20 ms, 20 Hz) was delivered to the implant for 4 s every 20 s, equivalent to the stimulation administered during a typical pup interaction trial. The position of the animal was tracked and analysed by Ethovision XT 8 software (Noldus) to calculate the average velocity and moved distance. For intruder assays, an 8–12-week-old C57BL/6J intruder of the opposite sex (receptive virgin female, as determined by vaginal smear, or sexually experienced male) was introduced into the resident mouse cage and 473-nm light was delivered in 20-ms pulses at 20 Hz for 1–4 s whenever the animal contacted the intruder with its snout. Sniffing and grooming durations were scored over a period of 5 min, aggression was scored during a 10-min period. After behavioural testing, animals were transfused transcardially and fibre placement as well as efficient light transmission were verified.

Statistics and reproducibility. Data were analysed by two-tailed, unpaired or paired Student's *t*-test, by two-tailed Fisher's exact test or by χ^2 test if not indicated otherwise, using Graph Pad Prism 7 for Mac OS, MATLAB or R. Statistical details are given in the respective figure legends. Experiments were independently performed twice (Figs. 1b–f, 2e–g, k, l, 3c–l, 4 and Extended Data Figs. 1, 2a–d, i, j, 3d, e, 4b–f, 7, 8), three (Figs. 1g–j, F2b, c, h, i, 3n–p and Extended Data Fig. 6b–d) or four times (Extended Data Fig. 6f–h).

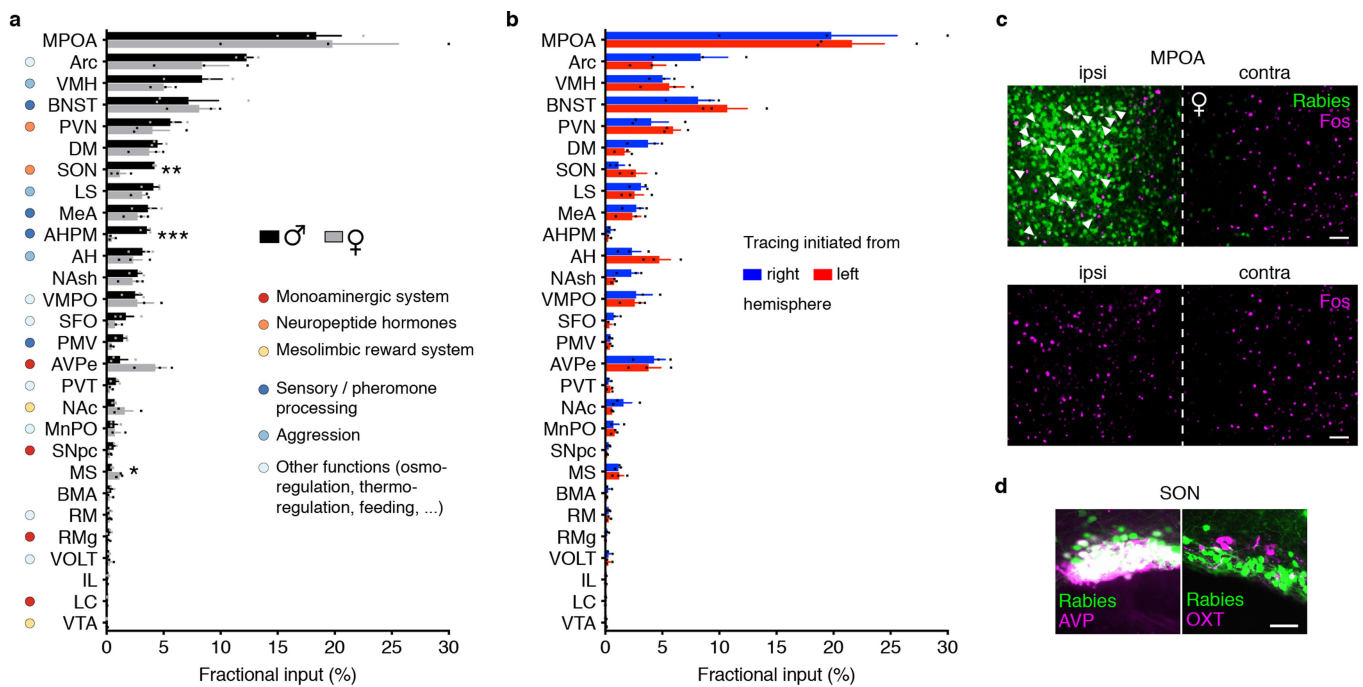
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The code that supports the findings of this study is available from the corresponding author upon request.

Data availability. The data that support the findings of this study are available from the corresponding author upon request.

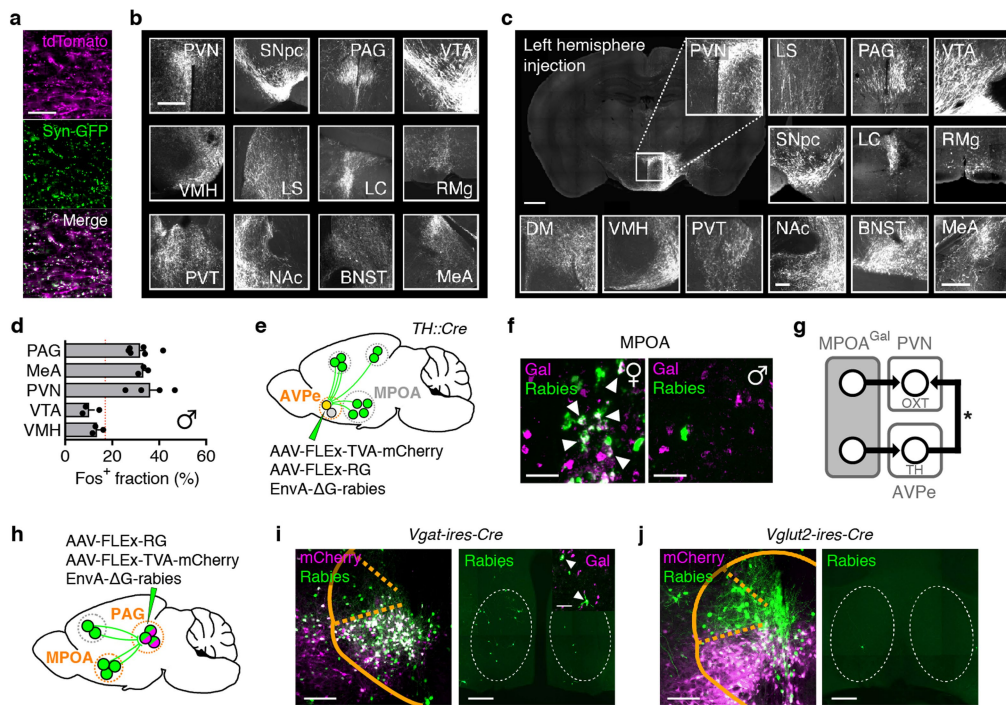
33. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
34. Miyamichi, K. et al. Dissecting local circuits: parvalbumin interneurons underlie broad feedback control of olfactory bulb output. *Neuron* **80**, 1232–1245 (2013).

35. Beier, K. T. et al. Circuit architecture of VTA dopamine neurons revealed by systematic input–output mapping. *Cell* **162**, 622–634 (2015).
36. Esposito, M. S., Capelli, P. & Arber, S. Brainstem nucleus MdV mediates skilled forelimb motor tasks. *Nature* **508**, 351–356 (2014).
37. Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd edn (Academic, New York, 2007).
38. Eliava, M. et al. A new population of parvocellular oxytocin neurons controlling magnocellular neuron activity and inflammatory pain processing. *Neuron* **89**, 1291–1304 (2016).
39. Xiao, L., Priest, M. F., Nasenbeny, J., Lu, T. & Kozorovitskiy, Y. Biased oxytocinergic modulation of midbrain dopamine systems. *Neuron* **95**, 368–384 (2017).
40. Swanson, L. W. & Sawchenko, P. E. Hypothalamic integration: organization of the paraventricular and supraoptic nuclei. *Annu. Rev. Neurosci.* **6**, 269–324 (1983).
41. Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat. Methods* **4**, 47–49 (2007).
42. Osakada, F. et al. New rabies virus variants for monitoring and manipulating activity and gene expression in defined neural circuits. *Neuron* **71**, 617–631 (2011).
43. Weible, A. P. et al. Transgenic targeting of recombinant rabies virus reveals monosynaptic connectivity of specific neurons. *J. Neurosci.* **30**, 16509–16513 (2010).
44. Menegas, W., Babayan, B. M., Uchida, N. & Watabe-Uchida, M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *eLife* **6**, e21886 (2017).
45. Sukikara, M. H., Mota-Ortiz, S. R., Baldo, M. V., Felicio, L. F. & Canteras, N. S. The periaqueductal gray and its potential role in maternal behavior inhibition in response to predatory threats. *Behav. Brain Res.* **209**, 226–233 (2010).
46. Fleming, A. S., Vaccarino, F. & Luebke, C. Amygdaloid inhibition of maternal behavior in the nulliparous female rat. *Physiol. Behav.* **25**, 731–743 (1980).
47. Numan, M., Numan, M. J. & English, J. B. Excitotoxic amino acid injections into the medial amygdala facilitate maternal behavior in virgin female rats. *Horm. Behav.* **27**, 56–81 (1993).
48. Sheehan, T., Paul, M., Amaral, E., Numan, M. J. & Numan, M. Evidence that the medial amygdala projects to the anterior/ventromedial hypothalamic nuclei to inhibit maternal behavior in rats. *Neuroscience* **106**, 341–356 (2001).
49. Insel, T. R. & Harbaugh, C. R. Lesions of the hypothalamic paraventricular nucleus disrupt the initiation of maternal behavior. *Physiol. Behav.* **45**, 1033–1041 (1989).
50. Numan, M. & Corodimas, K. P. The effects of paraventricular hypothalamic lesions on maternal behavior in rats. *Physiol. Behav.* **35**, 417–425 (1985).
51. Lee, G. & Gammie, S. C. GABA_A receptor signaling in the lateral septum regulates maternal aggression in mice. *Behav. Neurosci.* **123**, 1169–1177 (2009).
52. D'Anna, K. L. & Gammie, S. C. Activation of corticotropin-releasing factor receptor 2 in lateral septum negatively regulates maternal defense. *Behav. Neurosci.* **123**, 356–368 (2009).
53. Hansen, S., Harthorn, C., Wallin, E., Löfberg, L. & Svensson, K. Mesotelencephalic dopamine system and reproductive behavior in the female rat: effects of ventral tegmental 6-hydroxydopamine lesions on maternal and sexual responsiveness. *Behav. Neurosci.* **105**, 588–598 (1991).
54. Hansen, S. Maternal behavior of female rats with 6-OHDA lesions in the ventral striatum: characterization of the pup retrieval deficit. *Physiol. Behav.* **55**, 615–620 (1994).
55. Li, M. & Fleming, A. S. The nucleus accumbens shell is critical for normal expression of pup-retrieval in postpartum female rats. *Behav. Brain Res.* **145**, 99–111 (2003).
56. Keer, S. E. & Stern, J. M. Dopamine receptor blockade in the nucleus accumbens inhibits maternal retrieval and licking, but enhances nursing behavior in lactating rats. *Physiol. Behav.* **67**, 659–669 (1999).
57. Numan, M. et al. The effects of D1 or D2 dopamine receptor antagonism in the medial preoptic area, ventral pallidum, or nucleus accumbens on the maternal retrieval response and other aspects of maternal behavior in rats. *Behav. Neurosci.* **119**, 1588–1604 (2005).
58. Numan, M. & Nagle, D. S. Preoptic area and substantia nigra interact in the control of maternal behavior in the rat. *Behav. Neurosci.* **97**, 120–139 (1983).
59. Bridges, R. S., Mann, P. E. & Coppeta, J. S. Hypothalamic involvement in the regulation of maternal behaviour in the rat: inhibitory roles for the ventromedial hypothalamus and the dorsal/anterior hypothalamic areas. *J. Neuroendocrinol.* **11**, 259–266 (1999).
60. Numan, M. & Numan, M. A lesion and neuroanatomical tract-tracing analysis of the role of the bed nucleus of the stria terminalis in retrieval behavior and other aspects of maternal responsiveness in rats. *Dev. Psychobiol.* **29**, 23–51 (1996).
61. Numan, M., Rosenblatt, J. S. & Komisaruk, B. R. Medial preoptic area and onset of maternal behavior in the rat. *J. Comp. Physiol. Psychol.* **91**, 146–164 (1977).
62. Bridges, R. S., Numan, M., Ronsheim, P. M., Mann, P. E. & Lupini, C. E. Central prolactin infusions stimulate maternal behavior in steroid-treated, nulliparous female rats. *Proc. Natl Acad. Sci. USA* **87**, 8003–8007 (1990).
63. Numan, M. & Numan, M. J. Projection sites of medial preoptic area and ventral bed nucleus of the stria terminalis neurons that express Fos during maternal behavior in female rats. *J. Neuroendocrinol.* **9**, 369–384 (1997).
64. Lonstein, J. S., Simmons, D. A., Swann, J. M. & Stern, J. M. Forebrain expression of *c-fos* due to active maternal behaviour in lactating rats. *Neuroscience* **82**, 267–281 (1997).



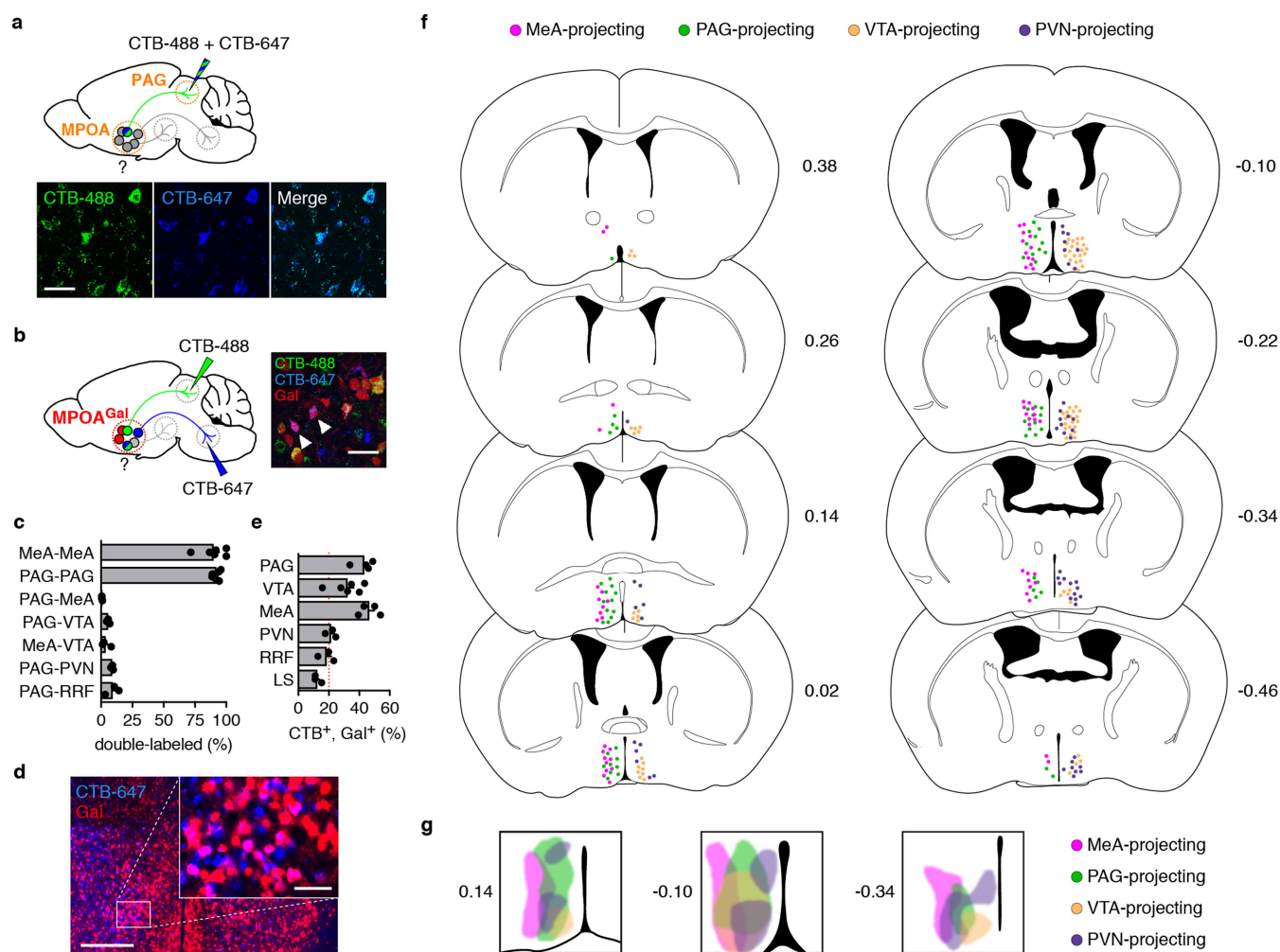
Extended Data Fig. 1 | Putative functional roles of brain areas providing monosynaptic inputs into MPOA^{Gal} neurons. **a**, Comparison between MPOA^{Gal} input fractions in virgin males ($n = 3$) and virgin females ($n = 3$) after rabies tracing (see Fig. 1a). Sexually dimorphic inputs are highlighted. Two-tailed t -tests, supraoptic nucleus (SON), $**P = 0.0041$; posteromedial amygdalo-hippocampal area (AHPM), $***P = 0.0007$; medial septum (MS), $*P = 0.0133$. **b**, Comparison between MPOA^{Gal} input fractions after rabies tracing was initiated from the right ($n = 3$) or left ($n = 3$) hemisphere in virgin females. No significant differences were found ($P > 0.05$; two-tailed paired t -test). **c**, Comparison between rabies-injected (ipsilateral (ipsi)) and non-injected (contralateral

(contra)) MPOA of a mother after parental behaviour. Activated (Fos⁺) rabies⁺ neurons are shown (top, arrowheads). Fos⁺ neuron numbers are not significantly different between hemispheres (bottom, $P = 0.43$, 95% confidence interval -4.176 – 1.843 ; two-tailed paired t -test; $n = 6$). **d**, MPOA^{Gal} neurons receive monosynaptic inputs from magnocellular SON^{AVP} neurons (mothers, $72.7 \pm 9.3\%$ overlap, $n = 3$; virgin females, $77.4 \pm 4.3\%$, $n = 3$; fathers, $83.3 \pm 3.3\%$, $n = 3$) but rarely from SON^{OXT} neurons (mothers, $4.6 \pm 4.2\%$ overlap, $n = 2$; virgin females, $4.5 \pm 1.0\%$, $n = 2$; fathers, $2.8 \pm 1.8\%$, $n = 2$). Data are mean \pm s.e.m. Scale bars, 100 μ m (c) and 50 μ m (d).



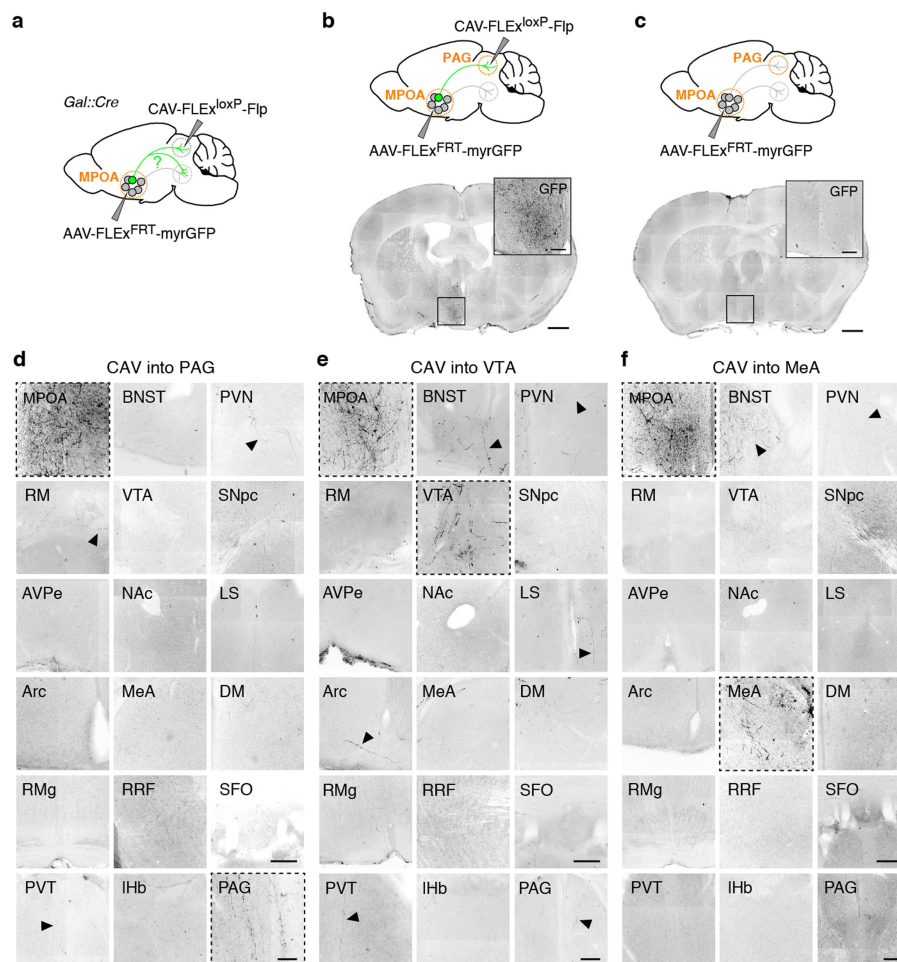
Extended Data Fig. 2 | MPOA^{Gal} projections in males and downstream connectivity. **a**, Synaptophysin–GFP (Syn–GFP) labelling of presynaptic sites in MPOA^{Gal} projections. **b**, Representative MPOA^{Gal} projections from a virgin male, identified by tdTomato fluorescence. **c**, Representative MPOA^{Gal} projections, identified by tdTomato fluorescence, after viral injection into the left MPOA. **d**, Fos⁺ fractions of virally labelled MPOA^{Gal} projections in fathers ($n = 6, 3, 4, 3, 3$, respectively, from top to bottom). Red line depicts the population average³. Data are mean \pm s.e.m. **e**, Trans-synaptic retrograde rabies tracing from AVPeTH neurons. **f**, MPOA^{Gal} neurons presynaptic to AVPeTH neurons in females (left, indicated by arrowheads, 21.4% Gal⁺ neurons, 47 out of 220 neurons, $n = 3$) and males

(right, 16.7% Gal⁺, 4 out of 24 neurons, $n = 2$). **g**, Direct and indirect MPOA^{Gal} to PVN^{OXT} connectivity. Asterisk, AVPeTH neurons form excitatory synapses with PVN^{OXT} neurons in females¹¹. **h**, Conditional monosynaptic retrograde tracing initiated from PAG. **i**, **j**, Injection sites with mCherry⁺ starter neurons in PAG of Vgat-IRES-Cre (**i**, left) or Vglut2-IRES-Cre (**j**, left) mice. Presynaptic, rabies⁺Gal⁺ neurons are detected in MPOA when tracing is initiated from PAG^{Vgat} (**i**, right, indicated by arrowheads), but not PAG^{Vglut2} (**j**, right), neurons. Scale bars, 50 μ m (**a**, **f** and **i**, inset), 200 μ m (**i** and **j**, left) 250 μ m (**b**, **c**, inset and **i** and **j**, right) and 500 μ m (**c**, left).



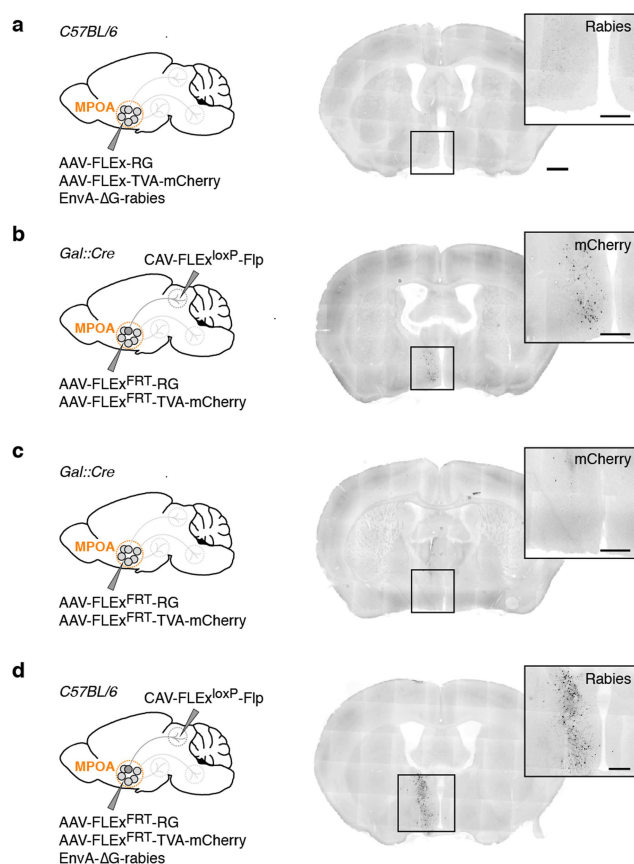
Extended Data Fig. 3 | MPOA^{Gal} projections correspond to mostly non-overlapping neuronal subpopulations. **a**, Control injection of a 1:1 mixture of CTB-488 and CTB-647 into PAG results in highly overlapping neuron populations in the MPOA (quantification in **c**). **b**, Strategy to determine collaterals between pairwise injected MPOA^{Gal} projections in *Gal::cre^{+/+}-loxP-Stop-loxP-tdTomato^{+/+}* mice. An example with two double-labelled MPOA^{Gal} neurons is shown after injection of CTB-488 into PAG and CTB-647 into VTA (right, indicated by arrowheads). **c**, Quantification of data in **a**, **b**. Data are mean \pm s.e.m. ($n = 6, 6, 3, 3, 3, 3, 3$, respectively, from top to bottom). **d**, Representative image from MPOA of *Gal::cre^{+/+}-loxP-Stop-loxP-tdTomato^{+/+}* mouse after injection of CTB-647 into PAG. Note high overlap between Gal⁺ and CTB⁺ neurons. **e**, Frequency of Gal⁺ neurons in individual, CTB-labelled MPOA

projections ($n = 4, 6, 4, 3, 3, 3, 3$, respectively, from top to bottom). Red line depicts expected labelling frequency, based on proportion of Gal⁺ MPOA neurons³ (around 20%). **c**, **e**, Data are mean \pm s.e.m. **f**, Distribution of cell bodies corresponding to specific MPOA^{Gal} projections. Individual MPOA^{Gal} projection areas in *Gal::Cre* virgin females were injected with Cre-dependent CAV2-FLEX-ZsGreen (see Fig. 2h). Only labelling patterns on the ipsilateral, injected side are shown and only two projection-specific subpopulations per side are displayed for clarity. Mouse brain images in this figure have been reproduced with permission from Elsevier³⁷. **g**, Zones occupied by MPOA^{Gal} cell bodies projecting to MeA, PAG, VTA and PVN in anterior (left), central (middle) and posterior (right) MPOA. **f**, **g**, Distance from bregma is shown in mm. Scale bars, 50 μ m (**a**, **b** and **d**, inset) and 250 μ m (**d**).

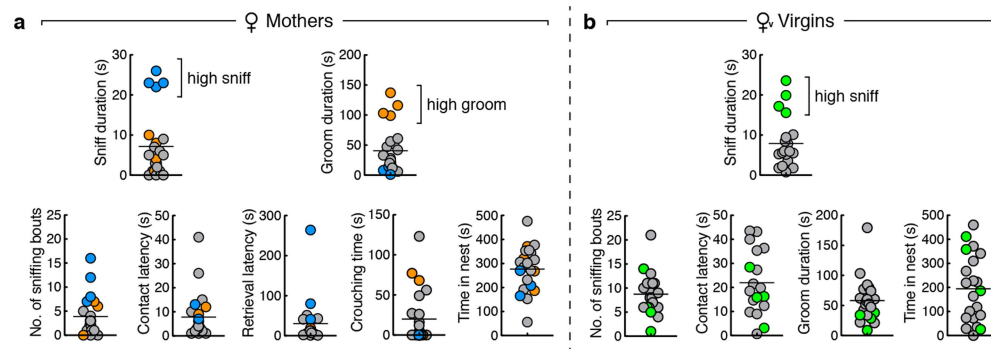


Extended Data Fig. 4 | MPOA^{Gal} projections barely collateralize.
a, Strategy to detect brain-wide axon collaterals of specific MPOA^{Gal} projections. **b**, Dense labelling of MPOA^{Gal} neurons after injection of retrograde tracer CAV into PAG and reporter AAV into MPOA. **c**, Absence of MPOA^{Gal} labelling in negative control without injection of CAV.

d–f, Only minor axon collaterals are detectable from MPOA^{Gal} neurons projecting to PAG (**d**; $n = 2$ virgin males), VTA (**e**; $n = 3$ virgin males) or MeA (**f**; $n = 2$ virgin males). Note the MPOA to MeA fibre tract in BNST in **f**. Signal was enhanced using anti-GFP immunostaining (Methods). Scale bars, **b**, **c**, 400 μm (**b**, **c**), 100 μm (insets) and 150 μm (**d–f**).

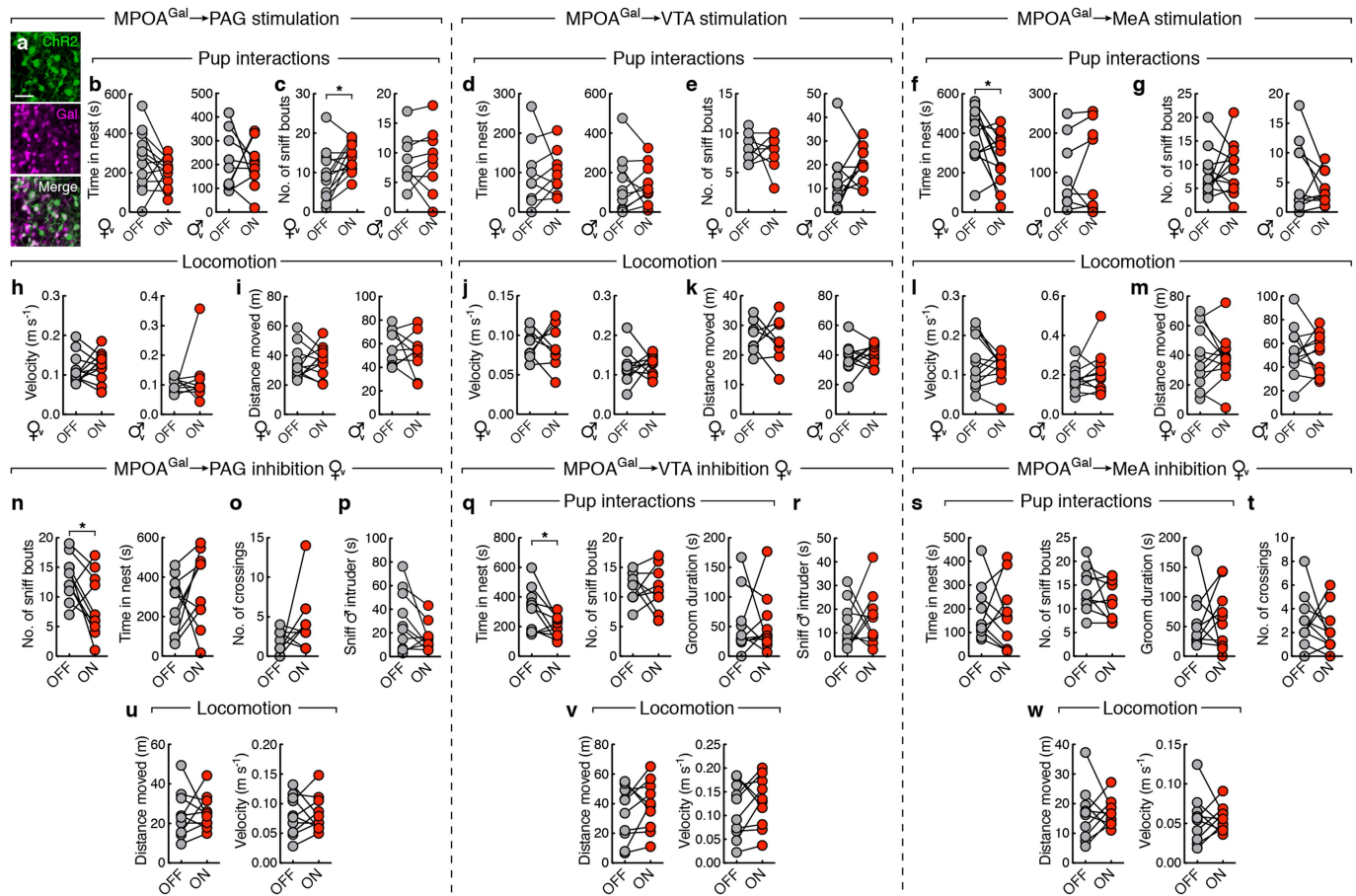


Extended Data Fig. 5 | Negative controls for monosynaptic retrograde tracing. **a**, Absence of rabies⁺ background labelling in the MPOA of AAV- and rabies-injected C57BL/6 control mice ($n = 2$). **b**, Labelling of MPOA^{Gal} neurons after injection of CAV into PAG and starter AAVs into MPOA of *Gal::cre* mice (261 ± 19 neurons, $n = 4$). **c**, Near-absence of labelling in AAV-only negative control (11 ± 2 neurons, $n = 2$). **d**, Background rabies⁺ neurons were present in the following brain areas of CAV-, AAV- and rabies-injected C57BL/6 control mice ($n = 3$): MPOA, BNST, anterior hypothalamus (AH), PVN and SON. These areas were therefore excluded from analysis (see Fig. 2k, l and Methods). Scale bars, 400 μ m (main images) and 150 μ m (insets).



Extended Data Fig. 7 | Distribution of parental behaviours in mothers and virgin females. Distribution of parental behaviours during 10-min pup interaction assays in mothers (**a**; $n = 23$) and virgin females (**b**; $n = 20$). In **a**, individuals exhibiting high pup sniffing are indicated

in blue across plots, and individuals exhibiting high pup grooming are indicated in orange. In **b**, individuals exhibiting high pup sniffing are indicated in green. Note that y axis ranges are identical between **a** and **b**. Lines depict mean.



Extended Data Fig. 8 | Behavioural specificity of MPOA^{Gal} projection stimulation. **a**, Channelrhodopsin-2 (ChR2) expression in MPOA^{Gal} neurons ($97.7 \pm 0.2\%$ overlap, virgin females, $n = 2$). Scale bar, $50 \mu\text{m}$. **b–g**, Effect of activating PAG (**b**, **c**), VTA (**d**, **e**) or MeA (**f**, **g**) projections on time spent in nest in virgin females and virgin males (**b**, $n = 13$ females and $n = 10$ males; **d**, $n = 9$ females and $n = 10$ males; **f**, $n = 10$ females and $n = 10$ males) and number of pup-directed sniffing bouts (**c**, $n = 13$ females and $n = 10$ males; **e**, $n = 9$ females and $n = 10$ males; **g**, $n = 10$ females and $n = 10$ males). **h–m**, Effect of activating PAG (**h**, **i**), VTA (**j**, **k**) or MeA (**l**, **m**) projections on locomotion velocity (**h**, $n = 13$ females and $n = 10$ males; **j**, $n = 8$ females and $n = 10$ males; **l**, $n = 10$ females and

$n = 10$ males) and moved distance (**i**, **k**, **m**). **n**, **q**, **s**, Effect of inhibiting PAG (**n**, $n = 10$ females), VTA (**q**, $n = 10$ females) or MeA (**s**, $n = 11$ females) projections on pup interactions. **o**, **t**, Effect of inhibiting PAG (**o**, $n = 10$ females) or MeA (**t**, $n = 11$ females) projections on number of barrier crosses. **p**, **r**, Effect of inhibiting PAG (**p**, $n = 10$ females) or MeA (**r**, $n = 11$ females) projections on chemoinvestigation of a male intruder. **u–w**, Effect of inhibiting PAG (**u**), VTA (**v**) or MeA (**w**) projections on locomotion velocity and moved distance ($n = 10$, 10 , 11 , respectively). Two-tailed paired t -tests; **c**, $*P = 0.0135$; **f**, $*P = 0.03$; **n**, $*P = 0.0413$, **q**: $*P = 0.0264$.

Extended Data Table 1 | List of brain areas and coordinates

| Abbreviation | Brain area | Injection coord. (AP / ML / DV) | Stimulation coord. (AP / ML / DV) | Recording coord. (AP / ML / DV) |
|--------------|---|------------------------------------|--------------------------------------|------------------------------------|
| AH | anterior hypothalamus | — | — | — |
| AHPM | posteromedial amygdalohippocampal area | — | — | — |
| Arc | arcuate nucleus | — | — | — |
| AVPe | anteroventral periventricular nucleus | 0.25 / 0.15 / -5.45 | — | — |
| BMA | basomedial amygdala | — | — | — |
| BNST | bed nucleus of the stria terminalis | — | — | — |
| DM | dorsomedial hypothalamus | — | — | — |
| IL | infralimbic cortex | — | — | — |
| LC | locus coeruleus | -5.4 / 0.88 / -2.65 | — | — |
| LS | lateral septum | 0.4 / 0.3 / -2.5 | — | — |
| MeA | medial amygdala | -1.6 / 2.25 / -4.95 | -1.6 / \pm 2.25 / -4.5 | — |
| MnPO | median preoptic nucleus | — | — | — |
| MPOA | medial preoptic area | 0 / 0.5 / -5.05 | — | 0 / 0.5 / -4.9 |
| MS | medial septum | — | — | — |
| NAc | nucleus accumbens - core | 1.0 / 0.7 / -3.8 | — | — |
| NAsh | nucleus accumbens - shell | — | — | — |
| PAG | (rostral) periaqueductal grey | -3.28 / 0.2 / -2.5 | -3.28 / \pm 0.2 / -2.2 | — |
| PeFA | perifornical area | -0.6 / 0.3 / -4.2 | — | — |
| PMV | ventral premammillary nucleus | — | — | — |
| PVN | periventricular hypothalamic nucleus | -0.82 / 0.25 / -4.6 | — | — |
| PVT | periventricular thalamic nucleus | -0.94 / 0 / -2.7 | — | — |
| RM | retromammillary nucleus | — | — | — |
| RRF | retrobulbar field | -4.04 / 1.0 / -3.4 | — | — |
| RMg | raphe magnus nucleus | -5.2 / 0 / -4.55 | — | — |
| SFO | subfornical organ | — | — | — |
| SNpc | substantia nigra pars compacta | -3.1 / 1.25 / -4.0 | — | — |
| SON | supraoptic nucleus | — | — | — |
| VMH | ventromedial hypothalamus | -1.5 / 0.4 / -5.7 | — | — |
| VOLT | vascular organ of the lamina terminalis | — | — | — |
| VTA | ventral tegmental area | -3.0 / 0.6 / -4.2 | -3.1 / \pm 0.5 / -4.1 | — |

Extended Data Table 2 | Summary of manipulations that affect parenting in MPOA^{Gal} target areas

| Brain area | Manipulation | Effect | Reference |
|------------|--|---|-----------------|
| PAG | Lesion | Facilitates maternal responses | 45 |
| | GABA _A receptor antagonist | Decreases maternal aggression, increases pup licking / grooming | 19 |
| MeA | Lesion | Accelerates onset of maternal behaviour | 46-48 |
| PVN | Lesion | Disrupts onset of maternal behaviour | 49 (but see 50) |
| LS | GABA _A receptor antagonist | Decreases maternal aggression | 51 |
| | Corticotropin releasing factor | Decreases maternal aggression | 52 |
| LC | Disruption of 5-HT production | Disrupts maternal behaviour (mice) | 53 |
| AVPe | Ablation of TH ⁺ neurons | Impairs maternal behaviour (mice) | 11 |
| | Optogenetic stimulation of TH ⁺ neurons | Enhances maternal behaviour (mice) | |
| VTA | Lesion | Impairs pup retrieval | 2,53 |
| | Inactivation | Impairs pup-paired conditioned place preference | 22 |
| NAc | Lesion | Impairs pup retrieval | 54,55 |
| | DA receptor antagonist | Inhibits retrieval and licking; enhances nursing | 56,57 |
| SNpc | Lesion | Disrupts maternal behaviour | 58 |
| VMH | Lesion | Accelerates onset of maternal behaviour | 59 |
| BNST | Lesion (ventral BNST) | Disrupts maternal behaviour | 60 |
| | Estrogen injection | Facilitates maternal responses | 61 |
| | Prolactin injection | Facilitates maternal responses | 62 |
| RRF | n/a | RRF-projecting MPOA neurons activated during maternal behaviour | 63 |
| PVT | n/a | Activated during maternal behaviour | 64 |

From those brain areas targeted by MPOA^{Gal} projections (Fig. 2c), manipulation of the following areas has been shown to affect maternal behaviour in rats (or mice where indicated)⁴⁵⁻⁶⁴. For a more comprehensive review see Kohl et al.¹⁰.

Innate immune memory in the brain shapes neurological disease hallmarks

Ann-Christin Wendeln^{1,2,3,12}, Karoline Degenhardt^{1,2,3,12}, Lalit Kaurani^{4,5}, Michael Gertig^{4,5}, Thomas Ulas⁶, Gaurav Jain^{5,7}, Jessica Wagner^{1,2,3}, Lisa M. Häsler^{1,2}, Katleen Wild^{1,2}, Angelos Skodras^{1,2}, Thomas Blank⁸, Ori Staszewski⁸, Moumita Datta⁸, Tonatiuh Pena Centeno^{5,7}, Vincenzo Capece^{5,7}, Md. Rezaul Islam⁵, Cemil Kerimoglu⁵, Matthias Staufenbiel^{1,2}, Joachim L. Schultze^{6,9}, Marc Beyer¹⁰, Marco Prinz^{8,11}, Mathias Jucker^{1,2}, André Fischer^{4,5} & Jonas J. Neher^{1,2*}

Innate immune memory is a vital mechanism of myeloid cell plasticity that occurs in response to environmental stimuli and alters subsequent immune responses. Two types of immunological imprinting can be distinguished—training and tolerance. These are epigenetically mediated and enhance or suppress subsequent inflammation, respectively. Whether immune memory occurs in tissue-resident macrophages in vivo and how it may affect pathology remains largely unknown. Here we demonstrate that peripherally applied inflammatory stimuli induce acute immune training and tolerance in the brain and lead to differential epigenetic reprogramming of brain-resident macrophages (microglia) that persists for at least six months. Strikingly, in a mouse model of Alzheimer's pathology, immune training exacerbates cerebral β -amyloidosis and immune tolerance alleviates it; similarly, peripheral immune stimulation modifies pathological features after stroke. Our results identify immune memory in the brain as an important modifier of neuropathology.

Contrary to the long-held assumption that immunological memory exists only in cells of the adaptive immune system, recent evidence has indicated that myeloid cells also display memory effects^{1,2}. For example, certain immune stimuli train blood monocytes to generate enhanced immune responses to subsequent immune insults^{3,4}. By contrast, other stimuli induce immune tolerance—suppression of inflammatory responses to subsequent stimuli^{3,5}. Innate immune memory lasts for several days in vitro and for up to three months in circulating monocytes in vivo and is mediated by epigenetic reprogramming in cultured cells, with chromatin changes also apparent in vivo^{3,6,7}. However, it is unclear whether immune memory occurs in long-lived tissue-resident macrophages and whether it alters tissue-specific pathology.

Microglia (brain-resident macrophages) are very long-lived cells^{8,9}. This makes them particularly interesting for studying immune memory, as virtually permanent modification of their molecular profile appears possible. As microglia are also involved in many neurological diseases^{10–12}, we investigated whether immune memory occurs in microglia in vivo and how it affects neuropathology.

Acute immune memory in the brain

It is well-established that inflammation in the periphery can prompt immune responses in the brain¹³. To evaluate whether immune memory can be induced in the brain by peripheral stimulation, we gave mice daily intraperitoneal injections of low-dose lipopolysaccharides (LPS) on four consecutive days, leading to mild sickness behaviour and temporary weight loss (Fig. 1a and Extended Data Fig. 1a). Three hours after the first LPS injection ($1 \times \text{LPS}$), there was a pronounced increase in blood cytokine levels, but only modest increases in brain cytokines. Upon the second injection ($2 \times \text{LPS}$), the blood levels of the pro-inflammatory cytokines IL-1 β , TNF, IL-6, IL-12 and IFN- γ

were diminished compared to their levels after $1 \times \text{LPS}$, whereas IL-10 release was not reduced, indicating peripheral immune tolerance. In sharp contrast, brain cytokines were markedly increased by $2 \times \text{LPS}$ injections, indicating a brain-specific training effect induced by the first LPS stimulus (Fig. 1b, c and Extended Data Fig. 2). Accordingly, a conspicuous morphological change in microglia occurred after $2 \times \text{LPS}$, whereas the number of activated (GFAP⁺) astrocytes increased only after $3 \times \text{LPS}$ (Extended Data Fig. 1b–d). Notably, $4 \times \text{LPS}$ virtually abolished TNF, IL-1 β and IL-6 release in the brain whereas IL-10 remained elevated, indicating immune tolerance.

Next, we examined the contribution of microglia to immune memory in the brain using inducible CX3CR1-CreER (Cre) mice crossed with mouse lines carrying *loxP*-flanked genes, in which tamoxifen-induced Cre expression results in persistent recombination in long-lived microglia but not in short-lived myeloid cells, including blood monocytes¹⁴. We induced microglial knockout of either transforming growth factor- β -activated kinase 1 (*Tak1*, also known as *Map3k7*), which results in inhibition of the NF- κ B, JNK and ERK1/2 pathways¹⁴, or histone deacetylases-1 and -2 (*Hdac1/2*), two major regulators of epigenetic reprogramming and macrophage inflammatory responses¹⁵. As expected, tamoxifen-induced knockout of either *Tak1* or *Hdac1/2* did not alter the peripheral inflammatory response. Furthermore, brain cytokine levels were indistinguishable after $1 \times \text{LPS}$, but the training effect following $2 \times \text{LPS}$ was virtually abolished in Cre⁺ mice. Notably, the cytokines showing the most pronounced training and tolerance effects (IL-1 β , TNF, IL-6) were also the most affected by microglial gene knockout (Fig. 1b, c and Extended Data Fig. 2), indicating that immune memory in the brain is predominantly mediated by microglia. Moreover, after $1 \times \text{LPS}$, Cre⁺ and Cre[−] mice showed indistinguishable weight loss (Extended Data Fig. 1a) and

¹German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany. ²Department of Cellular Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. ³Graduate School of Cellular and Molecular Neuroscience, University of Tübingen, Tübingen, Germany. ⁴Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. ⁵Department for Systems Medicine and Epigenetics in Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany. ⁶Genomics and Immunoregulation, LIMES-Institute, University of Bonn, Bonn, Germany. ⁷Bioinformatics Unit, German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany. ⁸Institute of Neuropathology, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ⁹Platform for Single Cell Genomics and Epigenomics at the University of Bonn and the German Center for Neurodegenerative Diseases, Bonn, Germany. ¹⁰Molecular Immunology in Neurodegeneration, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. ¹¹BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany. ¹²These authors contributed equally: Ann-Christin Wendeln, Karoline Degenhardt. *e-mail: jonas.neher@dzne.de

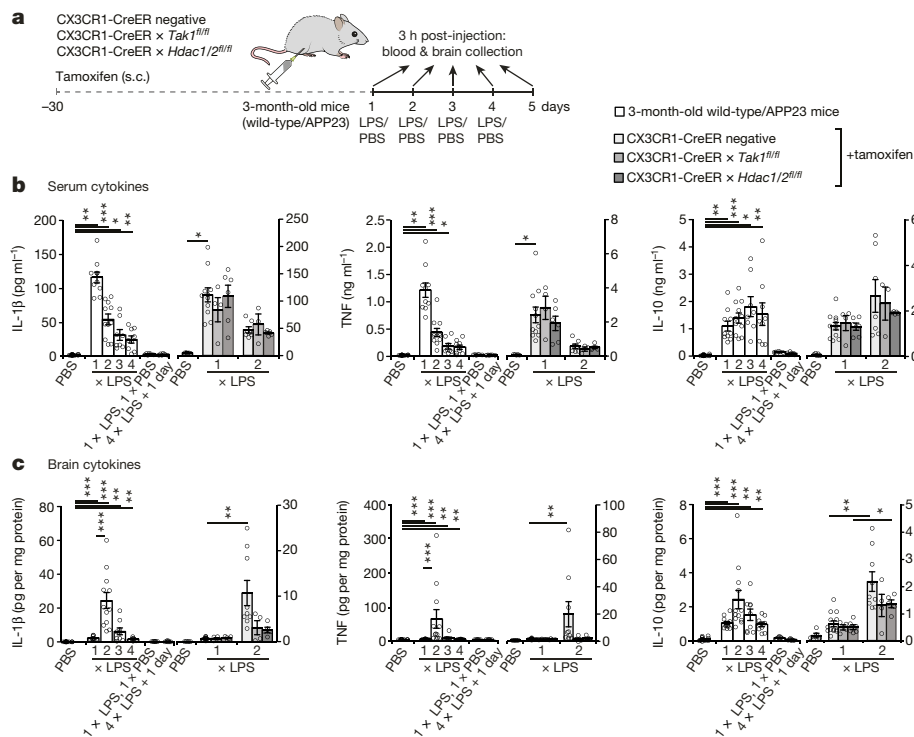


Fig. 1 | Peripheral immune stimulation evokes immune memory in microglia. **a**, Experimental approach. s.c., subcutaneous injection. **b**, Peripheral cytokine levels in wild-type or APP23 mice (white bars) and in mice with microglia-specific knockout of *Tak1* or *Hdac1/2* (coloured bars) following lipopolysaccharide (LPS) injections. Note that tolerance is induced with repeated injections. **c**, Brain cytokine levels, as in **b**. $2 \times$ LPS amplifies IL-1 β and TNF release in control mice, demonstrating immune

training; tolerance occurs after $3 \times$ LPS or $4 \times$ LPS. Cytokines return to baseline within 24 h ($1 \times$ LPS, $1 \times$ PBS and $4 \times$ LPS + 1 day). Microglia-specific knockout of *Tak1* or *Hdac1/2* selectively prevents immune training in the brain. In **b** and **c**, $n = 16, 11, 12, 9, 9, 7, 5, 13, 4, 6, 9, 4, 5$ from left to right. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ for independent-samples median test with correction for multiple comparisons. Data are means \pm s.e.m.

sickness behaviour (not shown); however, in animals with microglial *Tak1* knockout, sickness behaviour after $2 \times$ LPS was noticeably alleviated (Supplementary Video 1).

After intraperitoneal injections, LPS was found in the blood but not in the brain, indicating that at this dose neither significant entry of LPS into the brain nor opening of the blood–brain barrier occurred, as previously reported¹⁶. The latter was confirmed by the absence of blood iron in the brain parenchyma. Also, using type 2 CC chemokine receptor (CCR2) reporter mice¹⁷, we found no extravasation of circulating monocytes (Extended Data Fig. 1e–g), confirming that immune memory is mediated by brain-resident macrophages alone.

Immune memory shapes neuropathology

Next, we analysed whether the training- and tolerance-inducing stimuli ($1 \times$ LPS and $4 \times$ LPS, respectively) could lead to long-term alterations in brain immune responses and thereby modify disease pathogenesis. APP23 mice are a model of Alzheimer's disease pathology in which plaques of insoluble amyloid- β (A β) develop from 6 months of age. Amyloid plaques lead to activation of microglia¹⁸, thereby providing a stimulus that should reveal microglial immune memory. We injected 3-month-old APP23 mice with either $1 \times$ LPS or $4 \times$ LPS, then analysed pathology 6 months later (Fig. 2a). Strikingly, $1 \times$ LPS increased both plaque load and total A β levels compared to control animals, whereas $4 \times$ LPS decreased both plaque load and A β levels (Fig. 2b), with plaque-associated neuritic damage correlating directly with plaque size in all treatment groups (Extended Data Fig. 3a–c). In addition, the protein levels of A β precursor protein (APP) and its cleavage products were indistinguishable among the groups, indicating equivalent A β generation (Extended Data Fig. 3d). Furthermore, neither the total number of microglia nor the number of microglia clustering around plaques was altered by LPS treatments (Fig. 2c), whereas the

number of activated (GFAP⁺) astrocytes decreased slightly in animals that received either $1 \times$ LPS or $4 \times$ LPS injections (Extended Data Fig. 3e). However, the brain levels of IL-1 β , IL-6 and IL-12 were reduced in APP mice after $4 \times$ LPS, whereas IL-10 was reduced in APP mice after $1 \times$ LPS. By contrast, brain cytokine levels were not altered in wild-type littermate controls and baseline blood cytokine levels were unchanged in wild-type and APP animals. Furthermore, an additional LPS injection at 9 months of age caused peripheral cytokine responses that did not differ amongst LPS treatment groups in wild-type mice (Fig. 2d and Extended Data Fig. 4a–c). Thus, peripheral immune stimuli cause long-term alterations in the brain immune response and differentially affect Alzheimer's pathology.

To test for immune memory effects in a second disease model, we injected wild-type mice with $1 \times$ LPS or $4 \times$ LPS and induced focal brain ischaemia 1 month later. One day after ischaemia, neuronal damage and microglial numbers were indistinguishable amongst treatment groups (Fig. 3a), indicating that the initial ischaemic insult was unaffected by either $1 \times$ LPS or $4 \times$ LPS. However, the acute inflammatory response, which is driven by brain-resident cells early after ischaemia¹², differed, showing increased or decreased levels of IL-1 β in mice injected with $1 \times$ LPS or $4 \times$ LPS, respectively. By contrast, the release of IL-10 was suppressed only by $1 \times$ LPS (Fig. 3b), reminiscent of the results in APP mice (Fig. 2d). The levels of other brain cytokines and blood cytokines were indistinguishable amongst groups (Extended Data Fig. 5). Notably, 7 days after brain ischaemia, the volume of neuronal damage and microglial activation was strongly reduced in animals that received $4 \times$ LPS but unaffected by $1 \times$ LPS (Fig. 3c, d). These results confirm long-term modulation of brain immune responses and suggest persistent modification of stroke pathology following a tolerizing but not a training stimulus, possibly because the severity of the insult prevents its further exacerbation through amplification of the immune response.

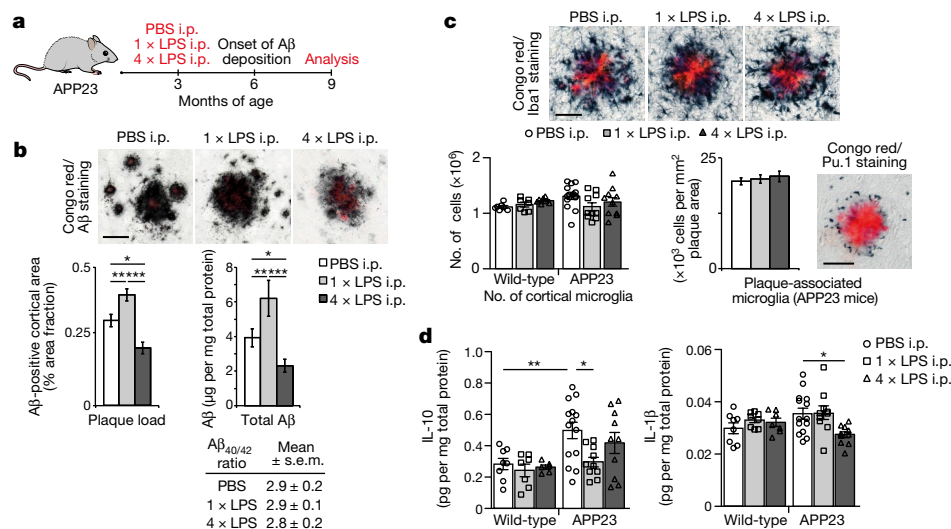


Fig. 2 | Cerebral β -amyloidosis is altered after peripheral immune stimulation. **a**, Experimental design. **b**, Analysis of cortical A β plaque load ($n = 22, 10, 10$ mice from left to right) and protein levels ($n = 14, 10, 10$ animals). **c**, **d**, Analysis of total cortical and plaque-associated microglia (**c**, $n = 7, 7, 7, 14, 10, 10$ mice) and cytokine levels of IL-10 and IL-1 β (**d**)

Microglial molecular profiles

In vitro, immune memory in macrophages results from epigenetically mediated alterations in the enhancer repertoire, leading to transcriptional changes^{3,19,20}. As our data indicate that acute immune memory in the brain is mediated predominantly by microglia, we isolated microglia by cell sorting (Extended Data Fig. 6) from 9-month-old animals stimulated with 1 × LPS or 4 × LPS at 3 months of age and performed chromatin immunoprecipitation for mono-methylation at lysine 4 of histone 3 (H3K4me1) and acetylation at lysine 27 of histone 3 (H3K27ac), which define active enhancers^{19,20}. Thus, we identified 20,241 putative active enhancers across all conditions.

First, we focused on H3K4me1 marks, which should mark all enhancers activated in response to the first and/or second immune stimulus (as enhancers may lose H3K27ac after cessation of inflammation but retain H3K4me1 marks)^{19,20}. Strikingly, H3K4me1 levels differed between control and LPS treatment groups both in wild-type and APP mice but also between mice treated with 1 × LPS or 4 × LPS (Extended Data Fig. 7b; Supplementary Table 1). For example, enhancers with increased H3K4me1 levels in microglia from wild-type

in wild-type and APP23 mice ($n = 8, 8, 7$ and $n = 14, 10, 10$ mice). i.p., intraperitoneal. Scale bar, 50 μ m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for one-way (**b**) and two-way ANOVA (**c**, **d**) with Tukey correction. Data are means \pm s.e.m.

mice injected with 1 × LPS versus 4 × LPS showed enrichment for the thyroid hormone signalling pathway, including a putative enhancer for hypoxia inducible factor-1 α (HIF-1 α). Similarly, enhancers with higher H3K4me1 levels in APP mice injected with 1 × LPS versus 4 × LPS were enriched for the HIF-1 signalling pathway. On the other hand, APP mice treated with 4 × LPS showed increased H3K4me1 levels in putative enhancers related to phagocytic function (Table 1a). Notably, we found no pathway enrichment when comparing H3K4me1 levels in microglia from APP and wild-type control mice (Table 1a), indicating that H3K4me1 levels were altered predominantly in response to LPS stimulation.

Next, we analysed enhancer activation by testing for differential regulation of H3K27ac levels. In line with the requirement of an acute stimulus for H3K27ac deposition¹⁹, differential enhancer activation was more pronounced in APP mice (where amyloid plaques activate microglia) than in wild-type mice (190 ± 18 in APP, 69 ± 5 in wild-type; Extended Data Fig. 7e; Supplementary Table 2). For example, differentially regulated H3K27ac levels in microglia from APP mice treated with 1 × LPS versus control APP mice were enriched for the

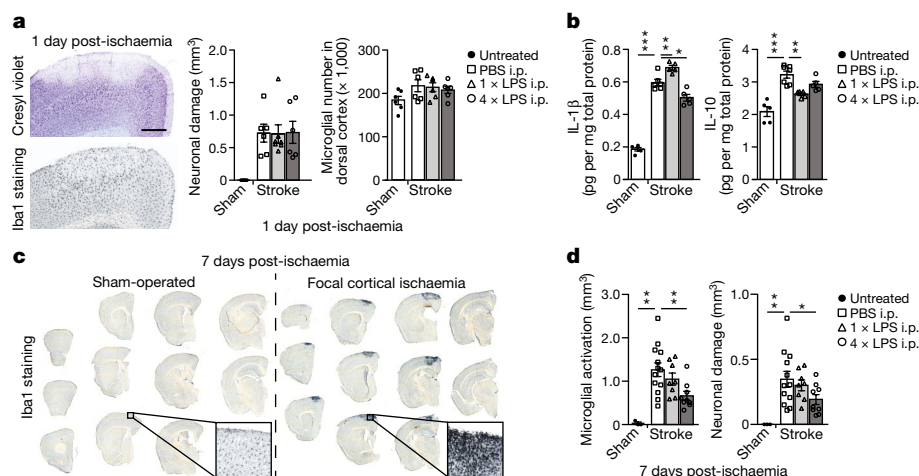


Fig. 3 | Stroke pathology is altered after peripheral immune stimulation. Pathological features of brain ischaemia induced one month after intraperitoneal (i.p.) injection with 1 × LPS or 4 × LPS. **a**, Neuronal damage (Cresyl violet, $n = 6, 6, 7, 6$ mice from left to right) and number of microglia (Iba1-positive, $n = 6, 6, 6, 6$ mice). **b**, Cytokine profiles 1 day

post-ischaemia ($n = 5, 7, 5, 5$ mice). **c**, **d**, Overview of microglial activation in the infarct (**c**) and quantification of neuronal damage and microglial activation (**d**) 7 days post-ischaemia ($n = 3, 13, 8, 9$ mice). Scale bar, 500 μ m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for one-way ANOVA with Tukey correction. Data are means \pm s.e.m.

Table 1 | The microglial enhancer repertoire 6 months after immune stimulation**a** KEGG pathway enrichment of differentially regulated H3K4me1 regions (threshold ≥ 1.5 -fold)

| | Condition 1 | Condition 2 | Increased in condition 1 | logP | Increased in condition 2 | logP |
|------------------|-------------|---------------|---|------|---------------------------------------|------|
| Wild-type | APP PBS | Wild-type PBS | | | | |
| | 1 × LPS | PBS | Renal cell carcinoma | −10 | | |
| | | | Focal adhesion | −6 | | |
| | | | MAPK signalling pathway | −5 | | |
| | | | Chemokine signalling pathway | −5 | | |
| | 4 × LPS | PBS | Endocytosis | −7 | PI3K-Akt signalling pathway | −4 |
| | | | Proteoglycans in cancer | −5 | | |
| | | | MAPK signalling pathway | −5 | | |
| | | | Colorectal cancer | −5 | | |
| | 1 × LPS | 4 × LPS | Proteoglycans in cancer | −6 | B cell receptor signalling pathway | −8 |
| | | | Thyroid hormone signalling pathway | −4 | MAPK signalling pathway | −8 |
| | | | | | FcγR-mediated phagocytosis | −8 |
| | | | | | Leishmaniasis | −7 |
| | | | | | Legionellosis | −5 |
| | | | | | Toll-like receptor signalling pathway | −5 |
| APP | 1 × LPS | PBS | Transcriptional misregulation in cancer | −8 | Focal adhesion | −5 |
| | | | Leukocyte transendothelial migration | −5 | TNF signalling pathway | −5 |
| | | | Glycosphingolipid biosynthesis | −5 | Rap1 signalling pathway | −5 |
| | 4 × LPS | PBS | Osteoclast differentiation | −11 | Osteoclast differentiation | −4 |
| | | | Leukocyte transendothelial migration | −9 | | |
| | | | Cytokine–cytokine receptor interaction | −6 | | |
| | | | Adherens junction | −6 | | |
| | | | FcγR-mediated phagocytosis | −6 | | |
| | | | Rap1 signalling pathway | −5 | | |
| | | | MAPK signalling pathway | −5 | | |
| | | | Endocytosis | −5 | | |
| | | | TGF-β signalling pathway | −4 | | |
| | | | Transcriptional misregulation in cancer | −4 | | |
| | 1 × LPS | 4 × LPS | Chemokine signalling pathway | −4 | | |
| | | | Salmonella infection | −6 | Rap1 signalling pathway | −10 |
| | | | Chagas disease | −5 | Osteoclast differentiation | −8 |
| | | | HIF-1 signalling pathway | −5 | Insulin signalling pathway | −8 |
| | | | Toxoplasmosis | −5 | Chemokine signalling pathway | −7 |
| | | | | | MAPK signalling pathway | −6 |
| | | | | | Glycosaminoglycan biosynthesis | −5 |

b KEGG pathway enrichment of differentially regulated H3K27ac regions (threshold ≥ 1.5 -fold)

| | Condition 1 | Condition 2 | Increased in condition 1 | logP | Increased in condition 2 | logP |
|------------------|-------------|---------------|---|------|---|------|
| Wild-type | APP PBS | Wild-type PBS | Thyroid hormone signalling pathway | −6 | | |
| | 1 × LPS | PBS | mTOR signalling pathway | −5 | | |
| | 4 × LPS | PBS | | | | |
| | 1 × LPS | 4 × LPS | Transcriptional misregulation in cancer | −5 | Transcriptional misregulation in cancer | −5 |
| APP | 1 × LPS | PBS | HIF-1 signalling pathway | −8 | Ras signalling pathway | −4 |
| | | | Thyroid hormone signalling pathway | −7 | MAPK signalling pathway | −7 |
| | | | Carbohydrate digestion and absorption | −6 | | |
| | | | Osteoclast differentiation | −5 | | |
| | | | AMPK signalling pathway | −5 | | |
| | | | Chronic myeloid leukaemia | −5 | | |
| | 4 × LPS | PBS | Rap1 signalling pathway | −5 | MAPK signalling pathway | −9 |
| | | | | | Osteoclast differentiation | −6 |
| | | | | | Thyroid hormone synthesis | −5 |
| | | | | | Chemokine signalling pathway | −4 |
| | 1 × LPS | 4 × LPS | Osteoclast differentiation | −12 | | |
| | | | Bacterial invasion of epithelial cells | −11 | | |
| | | | Toll-like receptor signalling pathway | −10 | | |
| | | | Ras signalling pathway | −9 | | |
| | | | Thyroid hormone signalling pathway | −9 | | |
| | | | FcγR-mediated phagocytosis | −9 | | |
| | | | mTOR signalling pathway | −8 | | |
| | | | Rap1 signalling pathway | −8 | | |
| | | | Regulation of actin cytoskeleton | −8 | | |
| | | | Carbohydrate digestion and absorption | −8 | | |
| | | | Phosphatidylinositol signalling system | −7 | | |
| | | | Chemokine signalling pathway | −7 | | |
| | | | Jak-STAT signalling pathway | −7 | | |
| | | | Focal adhesion | −7 | | |
| | | | Transcriptional misregulation in cancer | −6 | | |
| | | | Oestrogen signalling pathway | −6 | | |
| | | | TNF signalling pathway | −6 | | |
| | | | HIF-1 signalling pathway | −6 | | |
| | | | PI3K-Akt signalling pathway | −6 | | |
| | | | Chronic myeloid leukaemia | −6 | | |
| | | | Acute myeloid leukaemia | −5 | | |

Pathway enrichment of putative enhancers (with Benjamini–Hochberg correction) with differentially regulated H3K4me1 and H3K27ac levels (based on nearest gene; cumulative Poisson $P < 0.0001$). $n = 2$ replicates (8–10 mice per replicate).

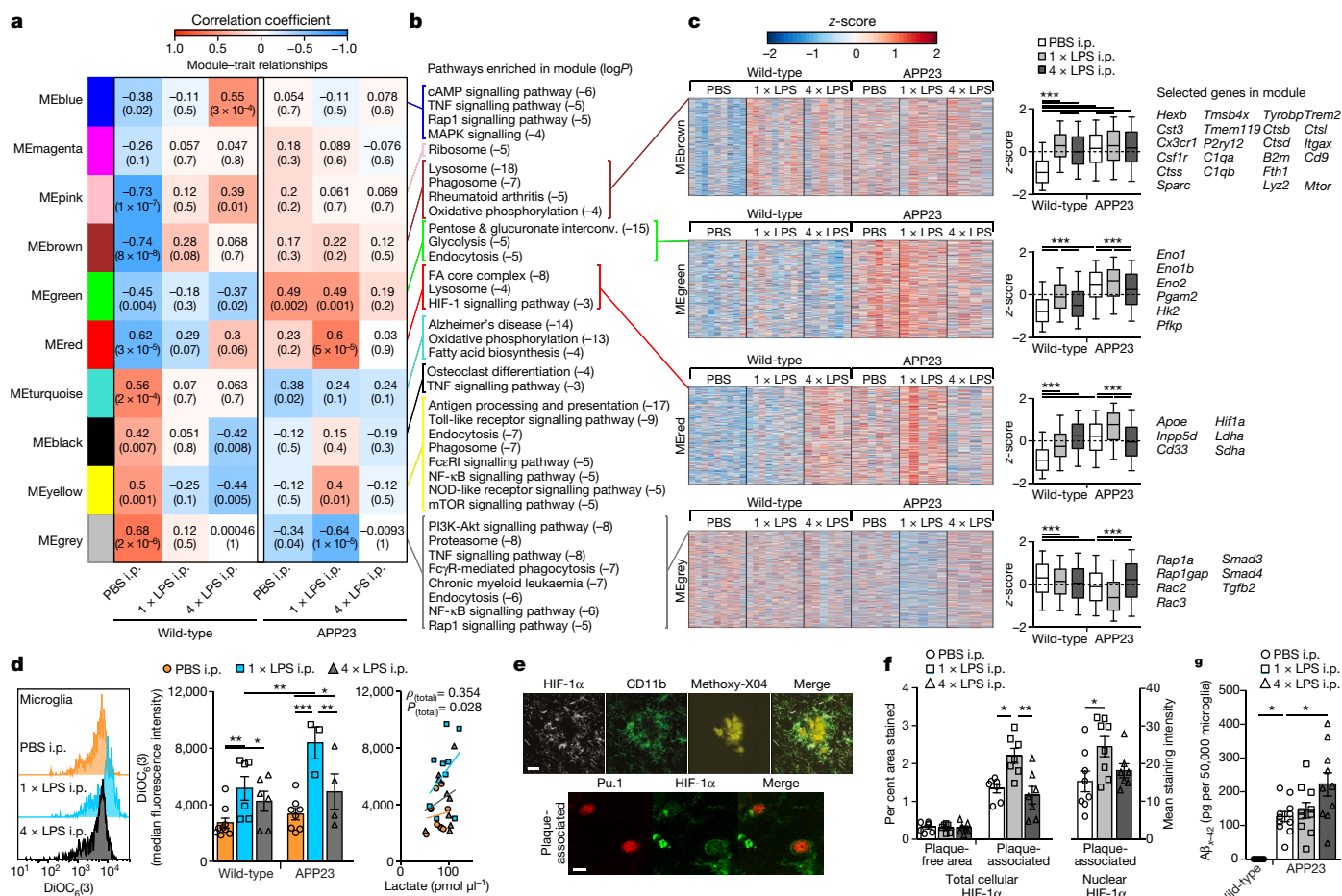


Fig. 4 | Microglial gene expression and function 6 months after immune stimulation. **a**, WGCNA (top, correlation coefficient; bottom, P value; $n = 9, 9, 6, 6, 5, 4$ mice from left to right). **b**, Selected KEGG pathways enriched in modules. **c**, Heatmaps of genes within modules, z-scores (boxplot whisker, 5th–95th percentile; $n = 1,601, 990, 949$ and $3,543$ genes in modules) and selected genes. **d**, Microglial mitochondrial membrane potential (left, middle; $n = 9, 6, 6, 8, 3, 4$ mice) and Pearson's correlation with lactate release (right;

$n = 11, 10, 10$ mice). **e**, Top, staining for HIF-1α, microglia (CD11b) and amyloid plaques (Methoxy-X04); bottom, staining for HIF-1α and microglial nuclei (Pu.1; single confocal plane) in brain sections from 9-month-old mice. Scale bars, 20 μm (top), 5 μm (bottom). **f**, Total cellular ($n = 7, 7, 7$ mice) and nuclear ($n = 8, 8, 7$ mice) HIF-1α staining intensity. **g**, Microglial Aβ content ($n = 5, 11, 10, 10$ mice). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for one-way ANOVA with Tukey correction. Data are means \pm s.e.m.

HIF-1 signalling pathway, with enhancer regions also being enriched for HIF-1α binding motifs (Table 1b and Extended Data Fig. 8), in line with changes in H3K4me1 levels (Table 1a) and the reported key role of HIF-1α in trained immunity and macrophage inflammatory responses^{4,21}.

Active enhancers in microglia from 4 × LPS-treated APP mice versus control APP mice showed enrichment only for the Rap1 signalling pathway, which has been implicated in phagocytosis of opsonized targets^{22,23}, again matching changes in H3K4me1 levels (Table 1). Strikingly, comparison of microglia from APP mice that received the training- (1 × LPS) or tolerance-inducing (4 × LPS) stimuli showed no pathway enrichment for active enhancers in mice injected with 4 × LPS, whereas enhancers in 1 × LPS-treated mice were enriched for a large number of inflammation-related pathways, highlighting the differential effects of the two immune memory states. Finally, comparison of microglia from vehicle-treated wild-type and APP mice demonstrated a small number of differentially activated enhancers with enrichment for the thyroid hormone signalling pathway (including a putative active enhancer for *Hif1a*) and the mTOR signalling pathway (Table 1b), indicating that microglia are also epigenetically reprogrammed in response to brain pathology alone.

We next examined microglial mRNA levels under the same conditions to determine whether epigenetic alterations were reflected in gene expression (Supplementary Table 3). First, we determined the concordance between 772 enhancers with significantly increased or decreased

H3K27ac levels (Supplementary Table 2) and the direction of change in the expression of their nearest gene. Indeed, there was a significant (albeit modest) concordance between alterations in H3K27ac levels and gene expression (median concordance of pairwise comparisons, 58%; $P = 0.03$). This suggested that gene expression is directly affected by the microglial active enhancer repertoire. Accordingly, weighted gene correlation network analysis (WGCNA²⁴) revealed striking parallels to epigenetic changes (Fig. 4a–c and Supplementary Table 4). For example, the red module (MERed; Fig. 4a) contained the *Hif1a* gene, showed enrichment for the HIF-1 signalling pathway and correlated strongly with the 1 × LPS-injected APP group. Furthermore, gene expression in MERed was upregulated in APP versus wild-type control mice and further increased by 1 × LPS, but downregulated by 4 × LPS.

HIF-1α activation in inflammatory-stimulated macrophages can occur downstream of mitochondrial hyperpolarization; enhanced HIF-1α signalling in turn promotes glycolysis, measurable as lactate release²⁵. Accordingly, the green module (MEgreen; Fig. 4a), which correlated positively with control and 1 × LPS-treated APP groups but negatively with control and 4 × LPS-treated wild-type groups, was found to be enriched in genes of the glycolysis pathway. Microglial gene expression in MEgreen was upregulated in APP versus wild-type control mice and further increased in APP mice by 1 × LPS but decreased in mice that received 4 × LPS. Therefore, we analysed mitochondrial membrane potential and lactate release in microglia. Strikingly, microglia from 1 × LPS-treated APP mice showed strongly increased

mitochondrial membrane potential, which correlated positively with the release of lactate (Fig. 4d), functionally corroborating the epigenetic and transcriptional alterations in trained microglia. Additionally, immunostaining confirmed an increase in protein levels of HIF-1 α in plaque-associated microglia; these levels were further increased in $1 \times$ LPS-treated APP mice (Fig. 4e, f). Thus, HIF-1 α signalling and a metabolic switch to glycolysis are activated in response to cerebral β -amyloid deposition, and are enhanced by immune training but reduced by immune tolerance in microglia.

In contrast to MEdred and MEgreen, the MEgrey module correlated positively with the control wild-type but negatively with the control APP and $1 \times$ LPS-treated APP groups. Compared to wild-type controls, microglial gene expression in MEgrey was downregulated in APP control animals and further decreased by $1 \times$ LPS, but showed unchanged levels in APP mice injected with $4 \times$ LPS compared with wild-type controls (Fig. 4a–c). Notably, MEgrey was enriched for phagocytosis-related pathways, including the Rap1 signalling pathway (Fig. 4a–c), again reflecting epigenetic changes (Table 1). We therefore tested whether phagocytosis of A β was enhanced in $4 \times$ LPS-treated APP mice. Indeed, microglial A β content was increased around 1.75-fold in these mice compared to APP control mice (Fig. 4g), providing further functional validation of the microglial enhancer repertoire and gene expression profiles.

Recent data have indicated that context-specific microglial phenotypes exist, for example, disease-associated microglia (DAM²⁶) and the microglial neurodegenerative phenotype (MGnD²⁷). Notably, the MEdbrown module, which was upregulated by both LPS treatments in wild-type mice as well as in all APP groups, contained a number of homeostatic microglial genes (for example, *Hexb*, *Cx3cr1* and *Csf1r*) but also all of the stage 1 DAM core genes except *ApoE*, as well as four of twelve stage 2 core genes²⁶ (Fig. 4c). Interestingly, the gene encoding ApoE, which may be crucial for promoting a detrimental microglial phenotype^{27,28}, was found in the same module (MEdred) as *Hif1a*. MEdred also contained other genes genetically linked to risk for Alzheimer's disease, namely *Cd33* and *Inpp5d*²⁹, suggesting that HIF-1 α may also be a detrimental modulator of Alzheimer's disease pathology.

The epigenetic landscape of microglia has been described only under homeostatic conditions^{30–32}. Our data now demonstrate epigenetic modifications in microglia in response to peripheral immune stimulation but also as a result of cerebral β -amyloidosis, including activation of the HIF-1 α and mTOR pathways, and leading to transcriptional and functional alterations. Although the global epigenetic and transcriptional changes were relatively modest, they are likely to have been driven by a small number of microglia that received the required secondary immune stimulation, as evidenced for example by increased levels of HIF-1 α in plaque-associated microglia (Fig. 4). mTOR activation is a well-known event in early Alzheimer's disease³³ and was recently shown in microglia, where it activated HIF-1 α and glycolysis to sustain microglial energy demand in models of Alzheimer's disease pathology³⁴. Our data now indicate that mTOR activation may be mediated by epigenetic microglial reprogramming in response to cerebral β -amyloidosis and that HIF-1 α signalling downstream of mTOR could be a detrimental event, because augmentation or suppression of HIF-1 α signalling occurred concomitantly with aggravated or alleviated A β deposition, respectively.

We here provide evidence of both immune training and tolerance in microglia and demonstrate their impact on neuropathology for the first time. While we cannot completely exclude the possibility that other cell-types contribute to immune memory and modulation of pathology in the brain, microglial-specific gene knockout of *Tak1* or *Hdac1/2* virtually abolished immune training (Fig. 1), indicating that microglia are likely to be the major effectors of immune memory. Notably, in our experiments, the effects of immune memory mostly became apparent following a secondary inflammatory stimulus, corroborating the concept of innate immune memory^{1,3}. However, while in the periphery training may be beneficial owing to enhanced pathogen elimination^{7,35,36}, and tolerance may be detrimental owing to higher

rates of infection resulting from immune suppression⁵, we found that training promotes, while tolerance alleviates, neuropathology. This is consistent with the beneficial effects of preventing microglial pro-inflammatory responses in models of Alzheimer's disease pathology and stroke^{12,37} and the worsening of cerebral β -amyloidosis in response to pro-inflammatory peripheral stimuli in animal models³⁸. Similarly, immune training has recently been described in epithelial stem cells, where it promotes wound healing but may also underlie autoimmune disorders³⁹. Thus, immune memory in the brain could conceivably affect the severity of any neurological disease that presents with an inflammatory component, but this will need to be studied for each individual condition.

Our data provide proof-of-principle for innate immune memory in microglia, and while our different LPS injection paradigms may not necessarily model physiological stimuli, we found that individual cytokines applied peripherally may also elicit immune memory effects in the brain (Extended Data Fig. 9). These results suggest that a wide variety of immune challenges may induce microglial immune memory and provide a possible mechanism for LPS-induced immune memory in the brain. It will be crucial to determine which other stimuli may lead to long-term modulation of microglial responses and thereby contribute to the severity of many neurological diseases.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0023-4>.

Received: 21 August 2016; Accepted: 23 February 2018;

Published online 11 April 2018.

1. Netea, M. G., Latz, E., Mills, K. H. G. & O'Neill, L. A. J. Innate immune memory: a paradigm shift in understanding host defense. *Nat. Immunol.* **16**, 675–679 (2015).
2. Netea, M. G. et al. Trained immunity: A program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).
3. Saeed, S. et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
4. Cheng, S. C. et al. mTOR- and HIF-1 α -mediated aerobic glycolysis as metabolic basis for trained immunity. *Science* **345**, 1250684 (2014).
5. Biswas, S. K. & Lopez-Collazo, E. Endotoxin tolerance: new mechanisms, molecules and clinical significance. *Trends Immunol.* **30**, 475–487 (2009).
6. Novakovic, B. et al. β -Glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* **167**, 1354–1368 (2016).
7. Kleinnijenhuis, J. et al. Bacille Calmette-Guérin induces NOD2-dependent non-specific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl Acad. Sci. USA* **109**, 17537–17542 (2012).
8. Tay, T. L. et al. A new fate mapping system reveals context-dependent random or clonal expansion of microglia. *Nat. Neurosci.* **20**, 793–803 (2017).
9. Füger, P. et al. Microglia turnover with aging and in an Alzheimer's model via long-term *in vivo* single-cell imaging. *Nat. Neurosci.* **20**, 1371–1376 (2017).
10. Prinz, M. & Priller, J. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* **15**, 300–312 (2014).
11. Heneka, M. T., Kummer, M. P. & Latz, E. Innate immune activation in neurodegenerative disease. *Nat. Rev. Immunol.* **14**, 463–477 (2014).
12. Iadecola, C. & Anrather, J. The immunology of stroke: from mechanisms to translation. *Nat. Med.* **17**, 796–808 (2011).
13. Perry, V. H., Cunningham, C. & Holmes, C. Systemic infections and inflammation affect chronic neurodegeneration. *Nat. Rev. Immunol.* **7**, 161–167 (2007).
14. Goldmann, T. et al. A new type of microglia gene targeting shows TAK1 to be pivotal in CNS autoimmune inflammation. *Nat. Neurosci.* **16**, 1618–1626 (2013).
15. Datta, M. et al. Histone deacetylases 1 and 2 regulate microglia function during development, homeostasis, and neurodegeneration in a context-dependent manner. *Immunity* [http://www.cell.com/immunity/fulltext/S1074-7613\(18\)30075-X](http://www.cell.com/immunity/fulltext/S1074-7613(18)30075-X) (2018).
16. Banks, W. A. & Robinson, S. M. Minimal penetration of lipopolysaccharide across the murine blood-brain barrier. *Brain Behav. Immun.* **24**, 102–109 (2010).
17. Saederup, N. et al. Selective chemokine receptor usage by central nervous system myeloid cells in CCR2-red fluorescent protein knock-in mice. *PLoS One* **5**, e13693 (2010).
18. Sturchler-Pierrat, C. et al. Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology. *Proc. Natl Acad. Sci. USA* **94**, 13287–13292 (1997).

19. Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
20. Kaikkonen, M. U. et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310–325 (2013).
21. Cramer, T. et al. HIF-1 α is essential for myeloid cell-mediated inflammation. *Cell* **112**, 645–657 (2003).
22. Caron, E., Self, A. J. & Hall, A. The GTPase Rap1 controls functional activation of macrophage integrin α M β 2 by LPS and other inflammatory mediators. *Curr. Biol.* **10**, 974–978 (2000).
23. Li, Y. et al. Rap1a null mice have altered myeloid cell functions suggesting distinct roles for the closely related Rap1a and 1b proteins. *J. Immunol.* **179**, 8322–8331 (2007).
24. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
25. Mills, E. L. et al. Succinate dehydrogenase supports metabolic repurposing of mitochondria to drive inflammatory macrophages. *Cell* **167**, 457–470 (2016).
26. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290 (2017).
27. Krasemann, S. et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**, 566–581 (2017).
28. Shi, Y. et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **549**, 523–527 (2017).
29. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
30. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
31. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
32. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **1617**, eaal3222 (2017).
33. Wang, C. et al. Targeting the mTOR signaling network for Alzheimer's disease therapy. *Mol. Neurobiol.* **49**, 120–135 (2014).
34. Ulland, T. K. et al. TREM2 maintains microglial metabolic fitness in Alzheimer's disease. *Cell* **170**, 649–663 (2017).
35. Kaufmann, E. et al. BCG educates hematopoietic stem cells to generate protective innate immunity against tuberculosis. *Cell* **172**, 176–190 (2018).
36. Arts, R. J. W. et al. BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host Microbe* **23**, 89–100 (2018).
37. Heneka, M. T., Golenbock, D. T. & Latz, E. Innate immunity in Alzheimer's disease. *Nat. Immunol.* **16**, 229–236 (2015).
38. O'Banion, M. K. Does peripheral inflammation contribute to Alzheimer disease? Evidence from animal models. *Neurology* **83**, 480–481 (2014).
39. Naik, S. et al. Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017).

Acknowledgements We thank P. Rizzu for experimental advice, L. Walker for manuscript comments and D. Bryce for statistical advice. This study was supported by a PhD fellowship from the Studienstiftung des Deutschen Volkes (A.-C.W.), a Roman Herzog Fellowship from the Hertie Foundation (J.J.N.), and grants from the network 'Neuroinflammation in Neurodegeneration' (State of Baden-Wuerttemberg, Germany; M.J. and M.P.), the Sobek-Stiftung (M.P.), the DFG (SFB992, Reinhart-Koselleck-Grant to M.P., SFB704 to J.L.S.), the European Research Council (A.F.) the Fortüne Program (Med. Faculty, Univ. Tuebingen; 2075-1-0; J.J.N.), the Fritz Thyssen Foundation (Cologne, Germany; J.J.N.) and the Paul G. Allen Family Foundation (Seattle, USA; J.J.N.). M.B. and J.L.S. are members of the Excellence Cluster ImmunoSensation.

Author contributions K.D., A.-C.W., J.W., L.M.H., K.W., A.S., T.B., O.S., M.D. and J.J.N. performed microglial isolation, in vivo and ex vivo experiments and histological/biochemical analyses. M.G., L.K., G.J., T.P.C., V.C., R.I., C.K., A.F., M.B., T.U., J.L.S. and J.J.N. performed ChIP-seq and RNA-seq analyses. J.J.N. conceived the study and coordinated the experiments together with M.J., A.F., M.P., M.B., J.L.S. and M.S. J.J.N. wrote the manuscript, with contributions from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0023-4>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0023-4>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.J.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. For all experiments, 3-month-old hemizygous APP23 transgenic (C57BL/6J-Tg(Thy1-APP_{K670N;M671L})23), APP23 transgene-negative littermate or C57BL/6J (wild-type) mice (Jackson Laboratory) or type 2 CC chemokine receptor (CCR2) reporter mice (kindly provided by R. Ransohoff, Boston) were used.

For experiments analysing immune responses after acute LPS and cytokine stimulation (see below), both male and female mice were used. For microglia-specific gene knockouts, CX3CR1-CreER animals were crossed with *Tak1*^{fl/fl} animals and Cre recombinase expression was induced by subcutaneous tamoxifen injections as previously described¹⁴. Similarly, microglial-specific knockout of *Hdac1/2* was achieved after crossing CX3CR1-CreER animals (kindly provided by S. Jung (Weizman Institute, Rehovot)) with a *Hdac1/2*^{fl/fl} line (kindly provided by P. Matthias (FMI Basel)). Both *Tak1*^{fl/fl} and *Hdac1/2*^{fl/fl} mice were injected with tamoxifen at 2–3 months of age and were incubated for four weeks without further treatment. Tamoxifen-injected CX3CR1-Cre-negative littermates were used as controls (because responses in CX3CR1-Cre-negative mice were indistinguishable in *Hdac1/2*^{fl/fl} and *Tak1*^{fl/fl} lines, pooled data are shown in Fig. 1).

As there is a significant gender effect on the pathology of both brain ischaemia and cerebral β -amyloidosis^{40,41}, only female mice were used for the analyses of brain pathology. APP23 mice express a transgene consisting of human APP with the KM670/671NL mutation under the Thy-1 promoter, and have been backcrossed with C57BL/6J mice for more than 20 generations. Female mice develop cerebral A β lesions in the neocortex at around 6 months of age¹⁸.

Animals were maintained under specific pathogen-free conditions. All experiments were performed in accordance with the veterinary office regulations of Baden-Württemberg (Germany) and were approved by the Ethical Commission for animal experimentation of Tübingen and Freiburg, Germany.

Peripheral immune stimulation. Three-month-old mice were randomly assigned to treatment groups and were injected intraperitoneally (i.p.) with bacterial lipopolysaccharides (LPS from *Salmonella enterica* serotype typhimurium, Sigma) at a daily dose of 500 μ g per kg bodyweight. Animals received either four LPS injections on four consecutive days (4 \times LPS), a single LPS injection followed by three vehicle injections on the following three days (1 \times LPS) or four vehicle injections (PBS). Acute stimulation showed indistinguishable cytokine responses in wild-type and APP23 transgenic animals; Fig. 1 shows the pooled data from both genotypes (see Extended Data Fig. 2 for data separated by genotype). Furthermore, as cytokine responses were indistinguishable in animals treated with one, two, three or four injections of PBS, pooled data from all time points are shown.

For peripheral cytokine treatments, recombinant mouse cytokines (TNF, IL-10; PeproTech) were aliquoted as per the manufacturer's instructions and stored at -80°C until use. To determine whether a long-term change in the brain's immune response (training or tolerance) occurred after peripheral cytokine injection, mice were treated on four consecutive days with 0.1 μ g per g bodyweight IL-10 or once with 0.1 or 0.2 μ g per g bodyweight TNF α . Control mice received four vehicle injections (PBS). Four weeks later, cytokine- and control-treated mice were injected with LPS (1 μ g per g bodyweight) or PBS, and were killed 3 h after the injection.

At the specified time-points, animals were deeply anaesthetized using sedaxylan and ketamine (64 mg/kg and 472 mg/kg), blood was collected from the right ventricle of the heart and animals were transcardially perfused with ice-cold PBS through the left ventricle. The brain was removed and sagittally separated into the two hemispheres, which were either fixed in 4% paraformaldehyde (PFA) or fresh-frozen on dry ice. Fresh-frozen hemispheres were homogenized using a Precellys lysing kit and machine at 10 or 20% (w/v) in homogenization buffer (50 mM Tris pH 8, 150 mM NaCl, 5 mM EDTA) containing phosphatase and protease inhibitors (Pierce). Fixed hemispheres were kept in 4% PFA for 24 h, followed by cryoprotection in 30% sucrose in PBS, subsequently frozen in 2-methylbutane and coronally sectioned at 25 μ m using a freezing-sliding microtome (Leica).

Focal brain ischaemia. For the induction of a focal cortical stroke, we modified existing models of endothelin-1 (ET-1)-induced brain ischaemia⁴² to avoid traumatic injury to the brain. Under anaesthesia and analgesia (fentanyl, midazolam and medetomidin: 0.05, 5 and 0.5 mg/kg bodyweight), 3-month-old mice were fixed in a stereotactic frame and a circular piece of skull was removed (5 mm diameter, centred on Bregma as described⁴³). The dura mater was carefully removed with the help of a microhook (Fine Science Tools) and 5 μ l ET-1 (Bachem; 64 μ M) in Hanks buffered salt solution (Invitrogen) or vehicle solution was topically applied to the cortex and incubated for 10 min. The craniotomy was then covered with a 5-mm glass coverslip, which was fixed in place with dental cement (Hybond), the skin was sutured, then the mice received antidote (flumazenil and atipamezol: 0.5 and 2.5 mg per kg bodyweight) and their health was monitored. Control mice underwent the same surgical procedure with application of vehicle solution to the cortex. After 4 weeks, animals were deeply anaesthetized and perfused as described above.

Western blotting analysis. For western blotting, total brain homogenates were sonicated three times for 5 s (LabSonic, B. Braun Biotech), and the protein levels of

the brain homogenates were quantified with a microplate bicinchoninic acid (BCA) assay (Pierce) and adjusted accordingly. Samples were then analysed on NuPage Bis-Tris gels (Invitrogen) using standard procedures. Proteins were transferred to nitrocellulose membranes, blocking was performed with 5% milk in PBS containing 0.05% Tween (PBST) for 1 h and blots were incubated with mouse anti-A β (6E10; 1:1,000, Covance) in PBST overnight at 4 $^{\circ}\text{C}$. Membranes were then probed with secondary HRP-labelled antibodies (1:20,000, Jackson ImmunoLaboratories). Protein bands were detected using chemiluminescent peroxidase substrate (ECL prime, GE Healthcare). Densitometric values of the protein band intensities were analysed with the software package Aida v.4.27 and normalized to GAPDH intensities.

Immunostaining. Immunohistochemical staining was performed on free-floating sections using either Vectastain Elite ABC kits (Vector laboratories) or fluorescent secondary antibodies (Jackson Immunolaboratories). Unless otherwise noted, brain sections were blocked for 1 h with 5% normal serum of the secondary antibody species, followed by primary antibody incubation overnight at 4 $^{\circ}\text{C}$. Primary antibodies used were: rabbit anti-Pu.1 (1:1,000, Cell Signalling), rabbit anti-Iba1 (1:1,000; Wako; catalogue no. 019-19741), rabbit anti-GFAP (1:500, Biozol; catalogue no. Z0334), rabbit anti-A β (CN3; 1:2,000)⁴⁴, mouse anti-HIF-1 α (1:500; Novus Biologicals, catalogue no. NB100-105, clone H1alpha67), rat anti-CD11b (1:2,000; Millipore, catalogue no. MAB1387Z), rabbit anti-APP (antibody 5313 to the ectodomain of APP, 1:750; kindly provided by C. Haass, DZNE Munich). Sections were then washed and incubated with secondary antibodies. Cresyl violet and Congo red staining was conducted according to standard procedures. Fluorescent plaque staining was achieved using Methoxy-X04 (4% vol of 10 mg/ml methoxy-X04 in DMSO, and 7.7% vol Cremophor EL in 88.3% vol PBS) for 20 min at room temperature.

Images were acquired on an Axioplan 2 microscope with Axioplan MRm and AxioVision 4.7 software (Carl Zeiss). Fluorescent images were acquired using a LSM 510 META (Axiovert 200 M) confocal microscope with an oil immersion 63 \times /1.4NA objective and LSM software 4.2 (Carl Zeiss), using sequential excitation of fluorophores. Maximum-intensity projections were generated using IMARIS 8.3.1 software (Bitmap).

For quantitative comparisons, sections from all groups were stained in parallel and analysed with the same microscope settings by an observer blinded to the treatment groups. To quantify the intensity of total microglial HIF-1 α staining, high-resolution bright-field images were acquired using fixed camera exposure time and lamp intensity and subsequently analysed with Fiji software. Colour channels were split and a fixed intensity threshold was applied to the red channel. On each image, the thresholded area over the total image area was calculated. Area fractions were measured on images of at least 9 plaques and 15 plaque-free regions per animal. To exclude an influence of plaque size on microglial HIF-1 α levels, plaques of similar size were selected for analysis of HIF-1 α levels in the different treatment groups (average plaque size: PBS i.p.: 1.73 ± 0.15 , 1 \times LPS i.p.: 1.84 ± 0.19 , 4 \times LPS i.p.: $2.27 \pm 0.39\%$ Congo red area fraction).

For nuclear HIF-1 α staining, a modified staining protocol was used. In brief, sections were blocked with mouse immunoglobulin blocking reagent (Vector laboratories) for 1 h at room temperature, followed by blocking with normal donkey serum for 1 h at room temperature. Sections were then incubated overnight with mouse anti-HIF1 α (clone mgc3, 1:50; Thermo Fisher Scientific, catalogue no. MA1-516) and rabbit anti-Pu.1 (1:250; New England Biolabs, catalogue no. 2258S, Clone 9G7) at 4 $^{\circ}\text{C}$. To quantify the intensity of nuclear HIF-1 α staining, z-stacks from three plaques and plaque-free regions per animal were acquired with the same microscope settings and subsequently analysed with IMARIS 8.3.1 software. Using the surfaces tool, a mask based on microglial nuclei was created using staining for Pu.1. A filter for area was applied to exclude background staining. The created surface was used to mask the HIF-1 α channel. The mean masked HIF-1 α intensity was then determined.

To quantify neuronal dystrophy, fluorescent images from 5–10 plaques per animal were acquired with the same microscope settings and subsequently analysed with Fiji software. Maximum intensity projections were generated to choose the region of interest consisting of APP staining and the plaque. Fluorescence channels were split, and intensity thresholds were applied to each channel. For every plaque, the thresholded area within the region of interest was calculated as a measure of plaque size and dystrophic area.

Stereological and morphological quantification. Stereological quantification was performed by a blinded observer on random sets of every 12th systematically sampled 25- μ m thick sections throughout the neocortex. Analysis was conducted using the Stereologer software (Stereo Investigator 6; MBF Bioscience) and a motorized x-y-z stage coupled to a video microscopy system (Optronics). For quantification of total Pu.1- and GFAP-positive cells, the optical fractionator technique was used with 3D disectors as previously described⁴⁵. For the quantification of plaque-associated cells, plaques were identified by Congo red staining and cells in their immediate vicinity were counted. Plaque load was determined by analysing the

cortical area covered by Congo Red and/or anti-A β staining using the area fraction fractionator technique⁴⁵. The volume of neuronal damage and microglial activation after brain ischaemia was determined using the Cavalieri estimator technique.

To analyse microglial morphology, we acquired three images from three non-consecutive Iba-1 immunostained brain sections per animal using identical camera acquisition settings, at $20\times/0.5\text{NA}$ magnification. In order to perform the filament tracing in IMARIS (v.8.3.1), images were pre-processed in Fiji to optimize their contrast for reconstruction. The image background was subtracted using the inbuilt Fiji plugin to obtain an evenly distributed intensity and enhance contrast to the cells; subsequently the images were sharpened and their intensity was adjusted to the respective minimum and maximum histogram values. Filaments were then traced in IMARIS using the in-built Autopath algorithm. Reconstruction parameters were kept constant among all images; each cell was reconstructed as a 'filament' element in IMARIS, associated with a total length and volume.

Enzyme-linked immunosorbent assay (ELISA). For quantification of A β by ELISA (Meso Scale Discovery) in brain homogenates or by single molecule array (SIMOA, Quanterix) in isolated microglial cells, samples were pre-treated with formic acid (Sigma-Aldrich, final concentration: 70% vol/vol), sonicated for 35 s on ice, and centrifuged at 25,000g for 1 h at 4°C. Neutralization buffer (1 M Tris base, 0.5 M Na₂HPO₄, 0.05% Na₂S₂O₃ (wt/vol)) was then added at a 1:20 ratio. A β was measured by an observer blinded to the treatment groups using human (6E10) A β triplex assay (Meso Scale Discovery, MSD) in brain homogenates or Simoa Human Abeta 42 2.0 Kit (Quanterix) in isolated microglia according to the manufacturer's instructions.

Soluble APP β containing the Swedish mutations (as present in the APP23 transgene) was measured using the sw soluble APP β kit (Mesoscale Discovery) following the manufacturer's instructions after extraction with 1% Triton X-100 and ultracentrifugation for 1 h (135,000g at 4°C).

For cytokine measurements, brain homogenates were centrifuged at 25,000g for 30 min at 4°C. Supernatants were analysed using the mouse pro-inflammatory panel 1 V-plex plate (Mesoscale Discovery) according to the manufacturer's instructions. To determine blood cytokines, serum was obtained by coagulation of whole blood in Vacuette (Greiner Bio-One) for 10 min at room temperature and centrifugation for 10 min at 2,000g. Serum samples were diluted 1:2 before measurements. The investigator was blinded to the treatment groups.

Measurements were performed on a Mesoscale Sector Imager 6000 or a Simoa HD-1 Analyzer. For analyses of brain homogenates, protein levels were normalized against total protein amount as measured by BCA protein assay (Pierce).

To determine levels of LPS in blood and brain homogenates, the Limulus Amebocyte Lysate assay was used according to the manufacturer's instructions (Pierce LAL Chromogenic Endotoxin Quantitation Kit). Standards were prepared either in serum or brain homogenate from non-injected control animals. Serum samples were diluted 1:100 and brain homogenates 1:5 to eliminate matrix effects. **Isolation of microglia and fluorescence-activated cell sorting analysis.** Fluorescence-activated cell sorting of microglia was performed on the basis of CD11b^{high} and CD45^{low} as previously described⁹ (Extended Data Fig. 6).

Assessment of microglial mitochondrial membrane potential and lactate release. To assess the microglial mitochondrial membrane potential, 10,000 microglia were sorted into 70 μ l PBS. Cells were incubated at 37°C with 3,3'-dihexyloxacarbocyanine iodide (DiOC₆(3); Thermo Fisher Scientific) at a final concentration of 0.2 nM for 20 min. At this concentration, mitochondrial dye accumulation is largely dependent on the mitochondrial membrane potential, with only minor contributions of the plasma membrane potential⁴⁶. After incubation, the cell suspension was diluted with ice-cold PBS and DiOC₆(3) fluorescence was immediately acquired with a Sony SH800 instrument.

For the assessment of microglial lactate release, 50,000 microglia from the same animals as used for DiOC₆(3) staining were plated in 96-well plates with 125 μ l macrophage serum-free medium (Thermo Fisher Scientific) and incubated for 24 h at 37°C with 5% CO₂. Lactate concentration in the medium was determined using a Lactate Assay Kit (BioVision) following the manufacturer's instructions and was correlated to DiOC₆(3) fluorescence values from cells of the same animal using IBM SPSS Statistics 22 software.

RNA sequencing. For RNA sequencing, 10,000 microglia were directly sorted into RNase-free PCR strips containing 30 μ l H₂O with 0.2% Triton X-100 and 0.8 U/ μ l RNase inhibitor (Clontech) and samples were immediately frozen on dry ice. RNA was isolated using NucleoSpin RNA XS kit (Macherey-Nagel) according to the manufacturer's instructions. Three nanograms total RNA was used as input material for cDNA synthesis. cDNA synthesis and enrichment were performed following the Smart-seq2 v4 protocol as described by the manufacturer (Clontech). Sequencing libraries were prepared with 1 ng of cDNA using the Nextera XT library preparation kit (Illumina) as described⁴⁷. Multiplexing of samples was achieved using three different index-primers in each lane. For sequencing, samples from each group (APP and wild-type) were pooled to rule out amplification and sequencing biases. Libraries were quality-controlled and quantified using a

Qubit 2.0 Fluorometer (Life Technologies) and Agilent 2100 Bioanalyzer (Agilent Technologies). A final library concentration of 2 nM was used for sequencing. Sequencing was performed using a 50-bp single read setup on the Illumina HiSeq 2000 platform.

Base calling from raw images and file conversion to fastq files were achieved by Illumina standard pipeline scripts (bcl2fastq v.2.18.0). Quality control was then performed using FASTQC (v.2.18.0) program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), eliminating one sample that had fewer than 20 million reads. Reads were trimmed off for sequencing adaptor and were mapped to mouse reference transcriptome (mm10) using STAR aligner 2.5.2b with non-default parameters. Unique read counts were obtained for each sample using HOMER v.4.8 software (<http://homer.salk.edu/homer/>) and the 'maketagdirectory -tbp 1' command, followed by 'analyzeRepeats.pl rna mm10 -count exons -noadj -condenseGenes'. Raw read counts were imported into R (v.3.2) and normalized using the Bioconductor (v.3.2) DESeq2 package (v.1.10.1) using default parameters. After normalization, all transcripts with a maximum overall group mean lower than 10 were removed. Unwanted or hidden sources of variation, such as batch and preparation date, were removed using the sva package⁴⁸. The normalized rlog transformed expression values were adjusted according to the surrogate variables identified by sva using the function removeBatchEffect from the limma package⁴⁹. To determine gene clusters associated with wild-type or APP23 animals following i.p. injections ($1\times\text{LPS}$ or $4\times\text{LPS}$) at 3 months of age, we then used the 13,627 present genes and applied the R implementation of WGCNA. We then performed WGCNA clustering using the '1-TOMsimilarityFromExpr' function with the network type 'signed hybrid', a power parameter of 7 (as established by scale-free topology network criteria), and a minimum module size of 50, dissecting the data into 10 modules. Finally, pathway enrichment analysis of genes within modules was performed using the 'findmotifs.pl' function of HOMER. Correction for multiple comparisons for KEGG pathway analyses was performed using the STATS package of R and applying the Benjamini-Hochberg correction. To focus on the most important molecular pathways, only pathways with $\log P \leq -3$ and at least five genes were considered.

Chromatin immunoprecipitation, library preparation and analysis. To isolate microglia for chromatin purification, 1 mM sodium butyrate, an inhibitor of histone deacetylases⁵⁰, was added to the dissection medium and FACS buffers. After staining, microglia were fixed in 1% PFA for 10 min at room temperature, followed by addition of glycine (final concentration: 125 mM) for 5 min and washing in HBSS. Microglia were then sorted into homogenization buffer (0.32 M sucrose, 5 mM CaCl₂, 5 mM magnesium acetate, 50 mM HEPES, 0.1 mM EDTA, 1 mM DTT, 0.1% vol/vol Triton X-100) and centrifuged at 950g for 5 min at 4°C. The pellet was resuspended in 100 μ l Nelson buffer (50 mM Tris, 150 mM NaCl, 20 mM EDTA, 1% vol/vol Triton X-100, 0.5% vol/vol NP-40) and frozen on dry ice.

Chromatin immunoprecipitation with sequencing (ChIP-seq) was performed as previously described⁵⁰, with slight modifications. In brief, two biological replicates were analysed for each condition and targeted histone modification. Cell lysates from 8–10 mice were pooled, giving a total cell number of approximately 0.8 million to 1 million cells per replicate. The cross-linked chromatin was sheared for 3×7 cycles (30 s on/off) in a BioruptorPlus (Diagenode) to achieve an average fragment size of 350 bp. Proper shearing and chromatin concentration were validated by DNA isolation and quantification using a small amount of each sample individually. Samples were split in half and 1 μ g of ChIP-grade antibody (H3K4me1: Abcam ab8895 or H3K27ac: Abcam ab4729) was added and incubated overnight at 4°C. From each sample, 1% of the total volume was taken as input control before antibody binding. Immunoprecipitation was performed by incubating samples with 30 μ l BSA-blocked protein A magnetic beads (Dynabeads, Invitrogen) for 1 h at 4°C. After purifying the precipitated chromatin and isolating the DNA, DNA libraries were generated using the Next Ultra DNA Library Prep Kit for Illumina and the Q5 polymerase (New England Biolabs). Multiplexing of samples was done using six different index primers from the Library Prep Kit. For each replicate, samples from each condition (genotype and treatment) were pooled to rule out amplification and sequencing biases within the final data. Input samples were pooled and processed accordingly. The ideal number of amplification cycles was estimated via RealTime PCR to avoid over-amplification. Accordingly, samples were amplified for 13–15 cycles and the DNA was isolated afterwards. Individual libraries were pooled; each pool represented one whole batch of samples for each condition and targeted histone modification and was set to a final DNA concentration of 2 nM before sequencing (50 bp) on a HiSeq 2000 (Illumina) according to the manufacturer's instructions.

Base calling from raw images and file conversion to fastq files were achieved using standard Illumina pipeline scripts. Sequencing reads were then mapped to the mouse reference genome (mm10) using rna-STAR aligner v2.3.0 with non-default parameters. Data were further processed using HOMER software (<http://homer.salk.edu/homer/>), following two recently published analyses on microglial epigenetic profiles^{30,31}. Tag directories were created from bam files using

'makeTagDirectory' for individual samples and inputs, and peak calling was performed using 'findpeaks -style histone' with fourfold enrichment over background and input, a Poisson P value of 0.0001, and a peak width of 500 bp for H3K4me1 and 250 bp for H3K27ac. Peaks common to both replicates were determined using 'mergepeaks' (-prefix) function. To focus analysis on enhancers, peaks within ± 2.5 kb of known transcription start sites were filtered out. Union peak files for H3K4me1 and H3K27ac marks were then created for group-wise comparisons using 'mergepeaks' function. Active enhancers, that is, genomic regions containing both H3K4me1 and H3K27ac peaks, were identified using the 'window' function of bedtools²⁵¹, requiring peaks of both marks to be located within a genomic region of 4 kb. Union peak files of active enhancers were then used for comparisons amongst groups for both H3K4me1 and H3K27ac marks using the 'getDifferentialPeaks' function (using a fold-change cut-off of 1.5 and a cumulative Poisson P value of 0.0001). Finally, differential peaks were annotated using the 'annotatepeaks.pl' function, including gene ontology analysis. Correction for multiple comparisons for KEGG pathway analyses was performed using the STATS package of R and applying the Benjamini–Hochberg correction. To focus on the most important molecular pathways, only pathways with $\log P \leq -3$ and at least three genes were considered.

For the generation of UCSC browser files, the 'makeUCSCfile' function was used, including normalization to respective input and library size, with a resolution of 10 bp. Files for heatmaps of 24 kb genomic regions and with a resolution of 250 bp were generated using the 'annotatePeaks.pl' function; clustering was then performed using Gene Cluster 3.0 and visualized using JavaTreeView 1.16r4.

To identify transcription factors involved in the differential activation of enhancers, the 'findMotifsGenome.pl' command was used to analyse a region of 500 bp around enhancer peaks ('-size 500'), as this resulted in more robust identification of motifs for known microglial lineage-determining transcription factors when determining motifs of all identified microglial enhancers (Extended Data Fig. 8). For all active enhancers, motif analysis was performed using the union H3K27ac peak file and standard background (that is, random genomic sequence created by HOMER). In the case of pairwise comparisons amongst conditions, the first condition's specific H3K27ac peak file was used as input and the second condition's peak file as background. Because motif enrichment was often relatively low, we focused on the most relevant results by determining transcription factor (families), whose motifs occurred at least twice in 'known' and 'de-novo' motifs.

Comparison between enhancer activation and gene expression. From our 14 pairwise comparisons (Fig. 4, Extended Data Fig. 7 and Supplementary Table 2), we analysed 772 differentially activated enhancers and compared increased or decreased H3K27ac levels with the direction of change in the expression of the nearest gene (difference in z -scores between the groups used for pairwise comparisons). The 14 concordance values were then statistically compared to chance level (50%) using a two-tailed Wilcoxon signed rank test.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

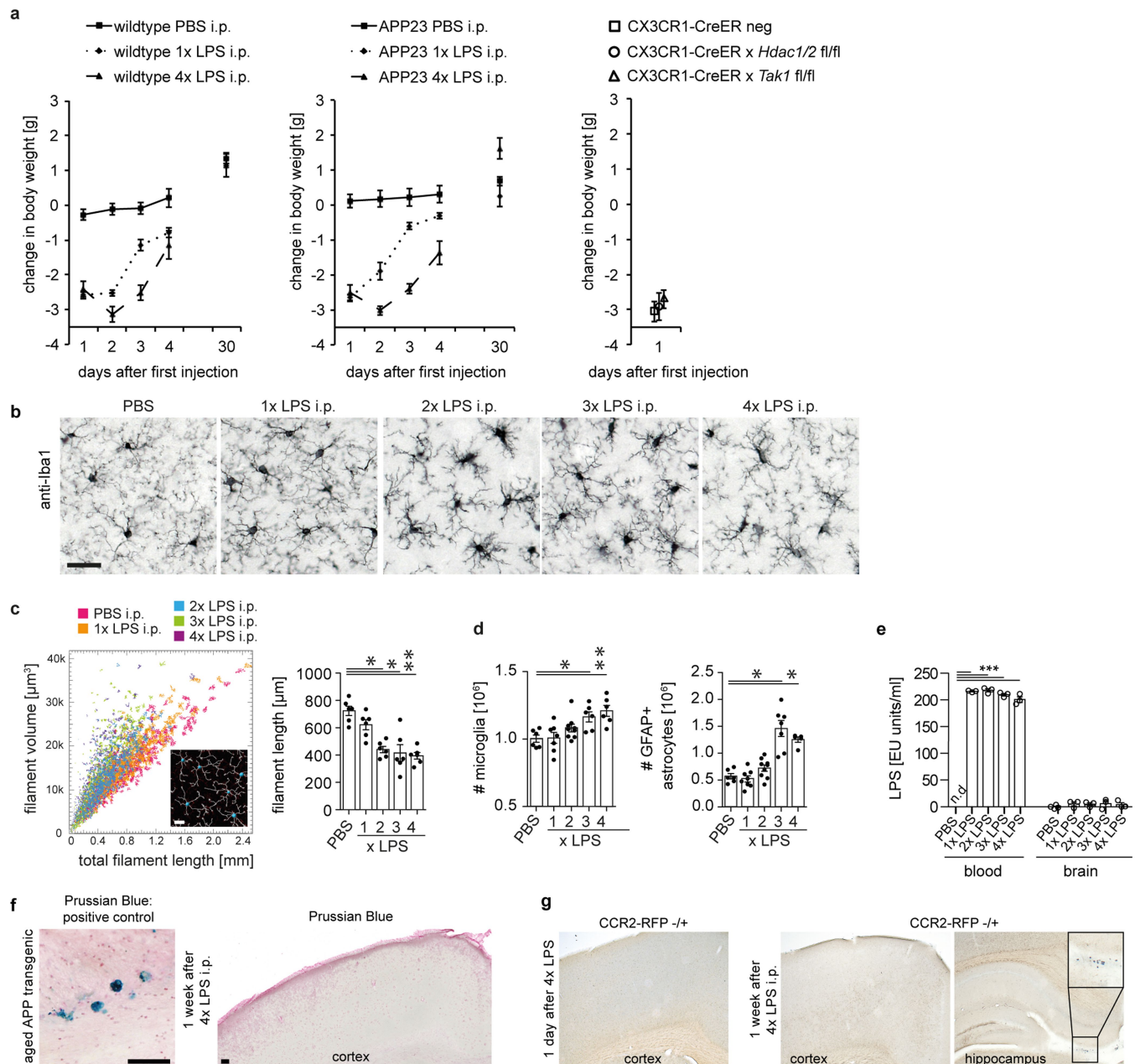
Statistics and reproducibility. Statistical analyses were performed using IBM SPSS Statistics 22 or Prism 5 software. Data were assessed for normal distribution (Shapiro–Wilk test) and statistical outliers using the 'explore' function. If the normality criterion was met, data were analysed using a one-way ANOVA (for

experiments on single genotypes), followed by pairwise comparison (if $P < 0.05$) with post-hoc Tukey correction (for samples with non-significant homogeneity of variance Levene's test) or Dunnett test (if homogeneity of variances not given). For comparisons across treatments and genotypes (for example, cytokine analyses in Fig. 2), a two-way ANOVA was performed, followed by post-hoc testing with Tukey correction for significant main effects ($P < 0.05$). As the cytokine data for acute LPS stimulation (Fig. 1) showed inequality of variance as well as skewedness, a non-parametric independent-samples median test was performed followed by pairwise comparison with correction for multiple comparison.

All experiments were performed at least twice and in independent batches of animals for key findings (figures show the pooled data). Owing to batch-related variation in some dependent variables, 'batch' was added as a random variable to analyses where a significant batch effect was observed. For datasets with small sample size (for example, western blotting analyses), the Kruskal–Wallis test was performed, followed by pairwise comparisons if $P < 0.05$. In the figure legends, n denotes the number of animals per treatment group. Minimum sample sizes were determined a priori using power analyses or as dictated by the methodology (for example, ChIP-seq).

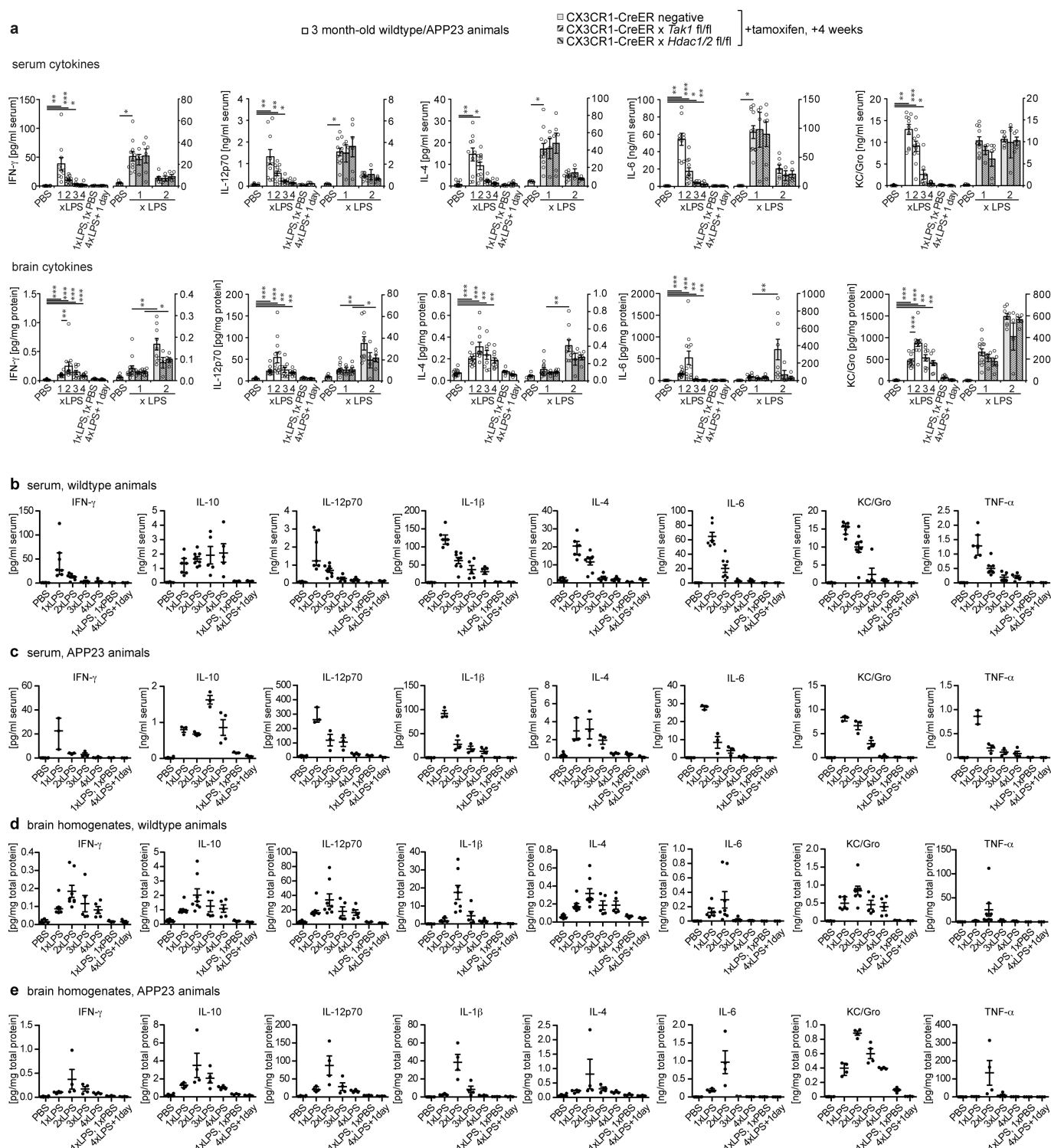
Raw and processed data are provided in the Gene Expression Omnibus (accession number GSE82170; subseries GSE82168 for ChIP-seq and GSE104630 for RNA-seq datasets). Other data that support the findings of this study are available from the corresponding author upon reasonable request.

40. Liu, F., Yuan, R., Benashski, S. E. & McCullough, L. D. Changes in experimental stroke outcome across the life span. *J. Cereb. Blood Flow Metab.* **29**, 792–802 (2009).
41. Sturchler-Pierrat, C. & Staufenbiel, M. Pathogenic mechanisms of Alzheimer's disease analyzed in the APP23 transgenic mouse model. *Ann. NY Acad. Sci.* **920**, 134–139 (2000).
42. Neher, J. J. et al. Phagocytosis executes delayed neuronal death after focal brain ischemia. *Proc. Natl Acad. Sci. USA* **110**, E4098–E4107 (2013).
43. Hefendehl, J. K. et al. Repeatable target localization for long-term *in vivo* imaging of mice with 2-photon microscopy. *J. Neurosci. Methods* **205**, 357–363 (2012).
44. Eisele, Y. S. et al. Peripherally applied A β -containing inoculates induce cerebral β -amyloidosis. *Science* **330**, 980–982 (2010).
45. Varvel, N. H. et al. Replacement of brain-resident myeloid cells does not alter cerebral amyloid- β deposition in mouse models of Alzheimer's disease. *J. Exp. Med.* **212**, 1803–1809 (2015).
46. Rottenberg, H. & Wu, S. Quantitative assay by flow cytometry of the mitochondrial membrane potential in intact cells. *Biochim. Biophys. Acta* **1404**, 393–404 (1998).
47. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
48. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
49. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
50. Halder, R. et al. DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nat. Neurosci.* **19**, 102–110 (2016).
51. Quinlan, A. R. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014).



Extended Data Fig. 1 | Acute responses to LPS injections. **a**, Weight changes after injection of LPS (wild-type mice: $n = 11, 11, 11, 11, 4$ from left to right for PBS, $n = 9, 9, 9, 8, 7$ for $1 \times$ LPS, $n = 10, 10, 10, 10, 7$ for $4 \times$ LPS; APP animals: $n = 14, 14, 14, 14, 7$ for PBS, $n = 8, 8, 8, 5, 5$ for $1 \times$ LPS; $n = 10, 10, 10, 10, 10$ for $4 \times$ LPS; Cre mice: $n = 5, 5, 4$). **b, c**, Morphological changes in microglia ($n = 6, 6, 6, 6, 6$ mice). Scale bar, $50 \mu\text{m}$. **d**, Numbers of microglia and activated (GFAP $^{+}$) astrocytes (microglia: $n = 6, 7, 8, 6, 6$ mice, astrocytes: $n = 6, 8, 9, 7, 5$ mice). **e**, Blood and brain levels of LPS after daily injections with $500 \mu\text{g}$ per kg bodyweight ($n = 4, 3, 3, 3, 3$ animals). **f**, Assessment of iron entry from

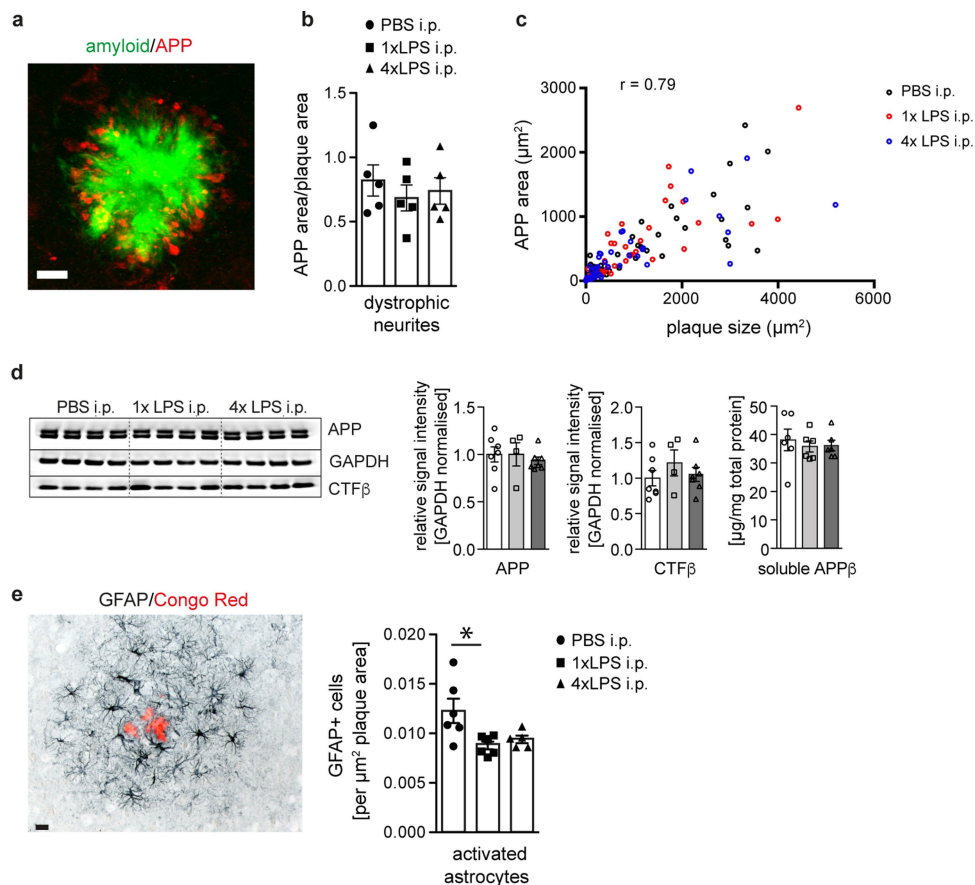
the blood (detected by Prussian blue staining) shows positive staining in an aged (>25 months) APP transgenic mouse, but not after repeated intraperitoneal LPS injections ($n = 3$ mice analysed). **g**, In heterozygous mice expressing red fluorescent protein (RFP) under the type 2 CC chemokine receptor (*Ccr2*) promoter, no entry of CCR2-expressing blood monocytes was detected after repeated LPS injection (staining for RFP; insert shows RFP-positive monocytes in the choroid plexus; $n = 3$ mice analysed). Scale bar, $100 \mu\text{m}$. Data are means \pm s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ for one-way ANOVA with Tukey correction.



Extended Data Fig. 2 | Cytokine response after acute LPS injections.

a. Additional cytokines (Fig. 1) analysed in the serum (top) and brain (bottom) 3 h after each daily intraperitoneal LPS injection on four consecutive days in 3-month-old mice (control mice received PBS injections; $n = 16, 11, 12, 9, 7, 7$ and $5, 13, 4, 6, 9, 4, 5$ mice for groups from left to right). **b, c.** Cytokine response in the blood only in wild-type

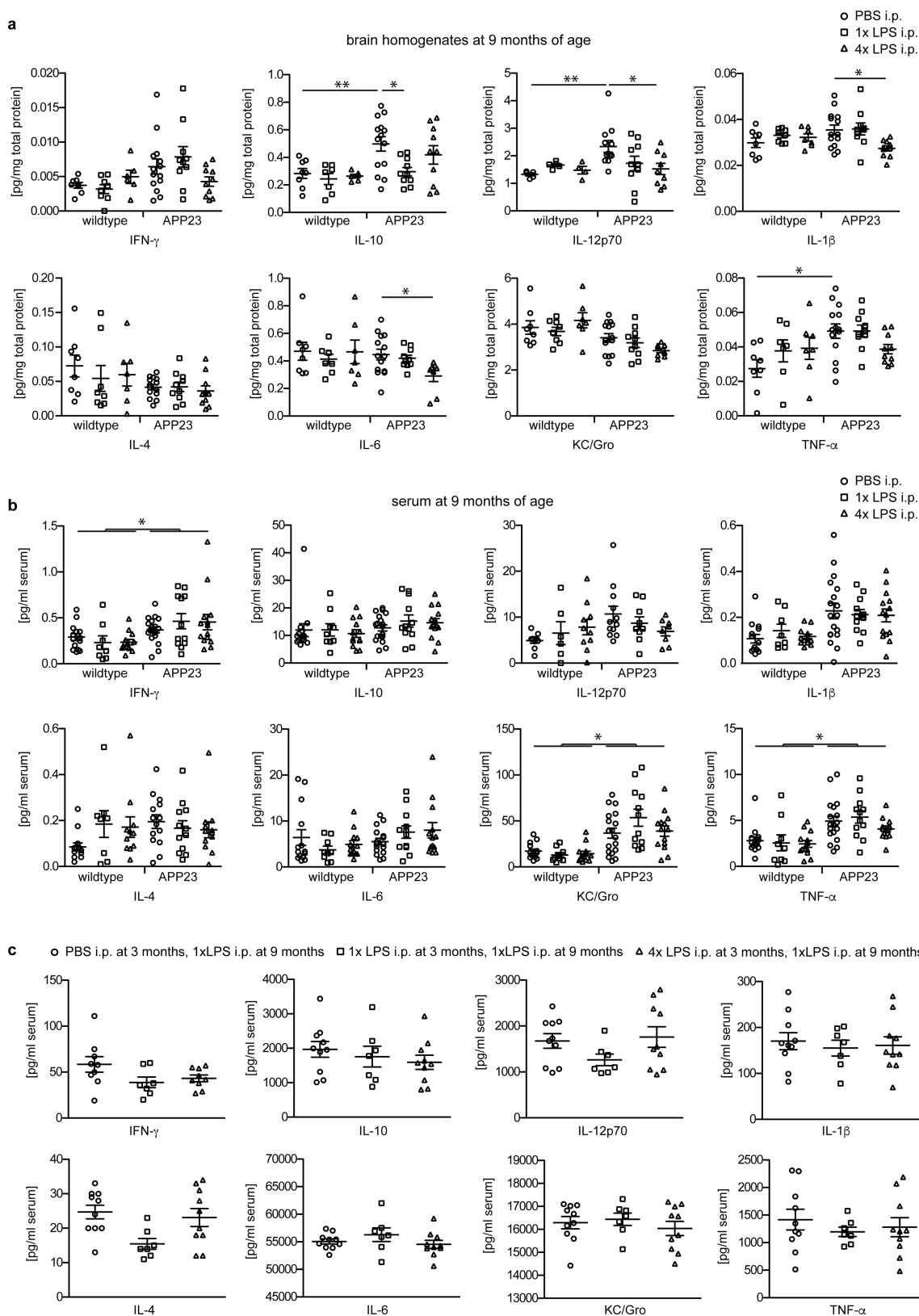
(**b**, $n = 6, 7, 8, 5, 5, 3, 3$ mice) or APP23 mice (**c**, $n = 10, 3, 3, 3, 4, 3, 3$ mice). **d, e.** Cytokine response in the brain only in wild-type (**d**, $n = 6, 7, 8, 5, 5, 3, 3$ mice) or APP23 mice (**e**, $n = 10, 4, 4, 4, 4, 4$ mice). Data are means \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for independent-samples median test with correction for multiple comparisons.



Extended Data Fig. 3 | APP levels and processing, neuritic dystrophy and astrocyte activation in 9-month-old APP23 animals.

a, b, Micrograph of fluorescent staining for amyloid plaque (Methoxy-X04; green) and amyloid precursor protein (APP; red) (**a**) shows neuritic dystrophy surrounding the amyloid deposit, which is unchanged by LPS treatments (**b**; $n = 5, 5, 5$ animals). **c**, Overall Pearson's correlation of plaque size with neuritic dystrophy (APP area; $n = 49, 39, 42$ plaques for PBS, $1 \times$ LPS, $4 \times$ LPS groups). **d**, Western blotting analysis (for gel

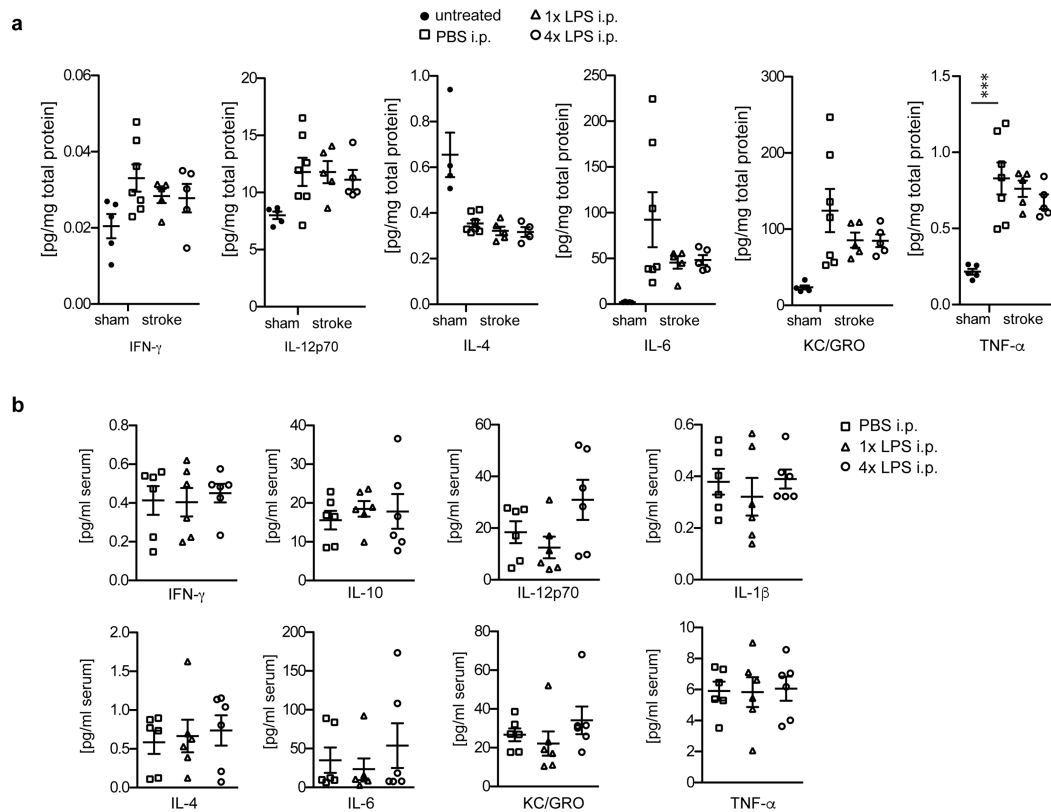
source data, see Supplementary Fig. 1) of brain homogenates for APP and C-terminal fragment- β (CTF β ; $n = 7, 4, 7$ mice), and soluble APP β ELISA ($n = 6, 6, 6$ mice). **e**, Micrograph of activated astrocytes (glial fibrillar acidic protein: GFAP) surrounding an amyloid plaque (Congo red) and quantification of the number of plaque-associated GFAP-positive astrocytes ($n = 6, 6, 5$ mice). Scale bar, $10 \mu\text{m}$ (**a**), $20 \mu\text{m}$ (**e**). Data are means \pm s.e.m. * $P < 0.05$ for one-way ANOVA with Tukey correction.



Extended Data Fig. 4 | Cytokine levels in 9-month-old animals.

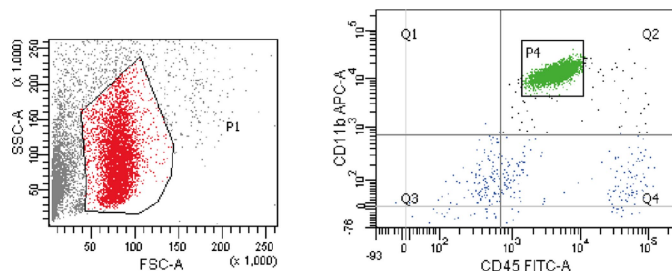
a, Cytokine measurements in brain homogenates of 9-month-old wild-type ($n = 8, 8, 7$ mice) and APP23 mice ($n = 14, 10, 10$ mice) treated i.p. with 1 × LPS or 4 × LPS at 3 months of age. **b**, Cytokine measurements in the serum of 9-month-old wild-type (WT; $n = 14, 9, 13$ mice) and APP23 mice (APP; $n = 18, 12, 14$ mice) after i.p. stimulation with 1 × LPS or 4 × LPS at 3 months of age. **c**, Cytokine measurements in the serum of

wild-type mice stimulated i.p. with 1 × LPS or 4 × LPS at 3 months of age and re-stimulated with an additional LPS injection ($500 \mu\text{g kg}^{-1}$) at 9 months of age ($n = 10, 7, 10$ animals). Data are means \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ for two-way ANOVA with Tukey correction. In **b** a significant main effect for genotype is indicated by bars spanning all conditions of the same genotype.

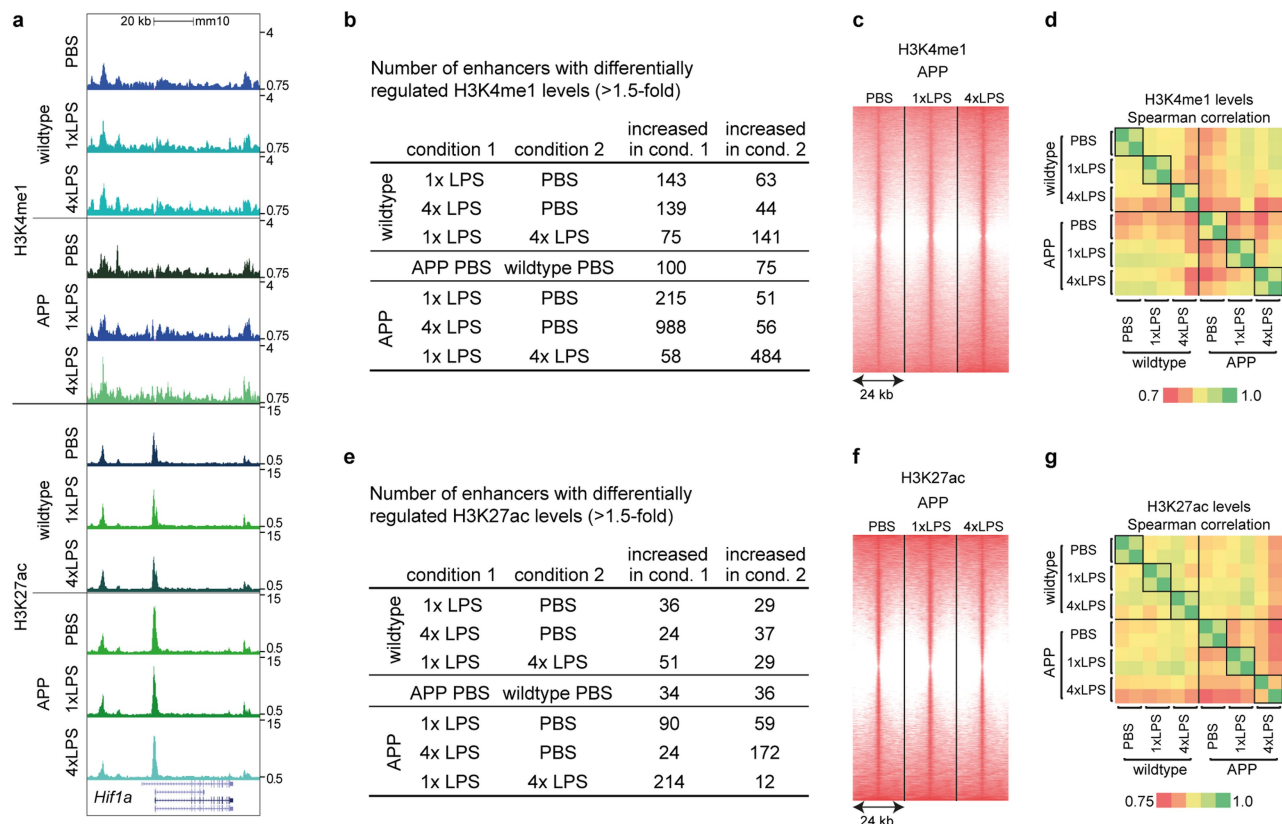


Extended Data Fig. 5 | Cytokine levels after brain ischaemia and in blood of 4-month-old mice. Three-month-old animals were injected i.p. with 1 \times LPS or 4 \times LPS and incubated for 4 weeks before receiving a stroke. **a**, Cytokine measurements in brain homogenates 24 h after stroke

($n = 5, 7, 5, 5$ animals). **b**, Cytokine measurements in the serum ($n = 6, 6, 6$ animals). Data are means \pm s.e.m. *** $P < 0.001$ for one-way ANOVA with Tukey correction.



Extended Data Fig. 6 | Microglial sorting strategy. Microglia were sorted as CD11b^{high} and CD45^{low} cells (population P4) from 9-month-old APP23 mice or wild-type littermates following i.p. injections of $1 \times$ LPS or $4 \times$ LPS at 3 months of age.



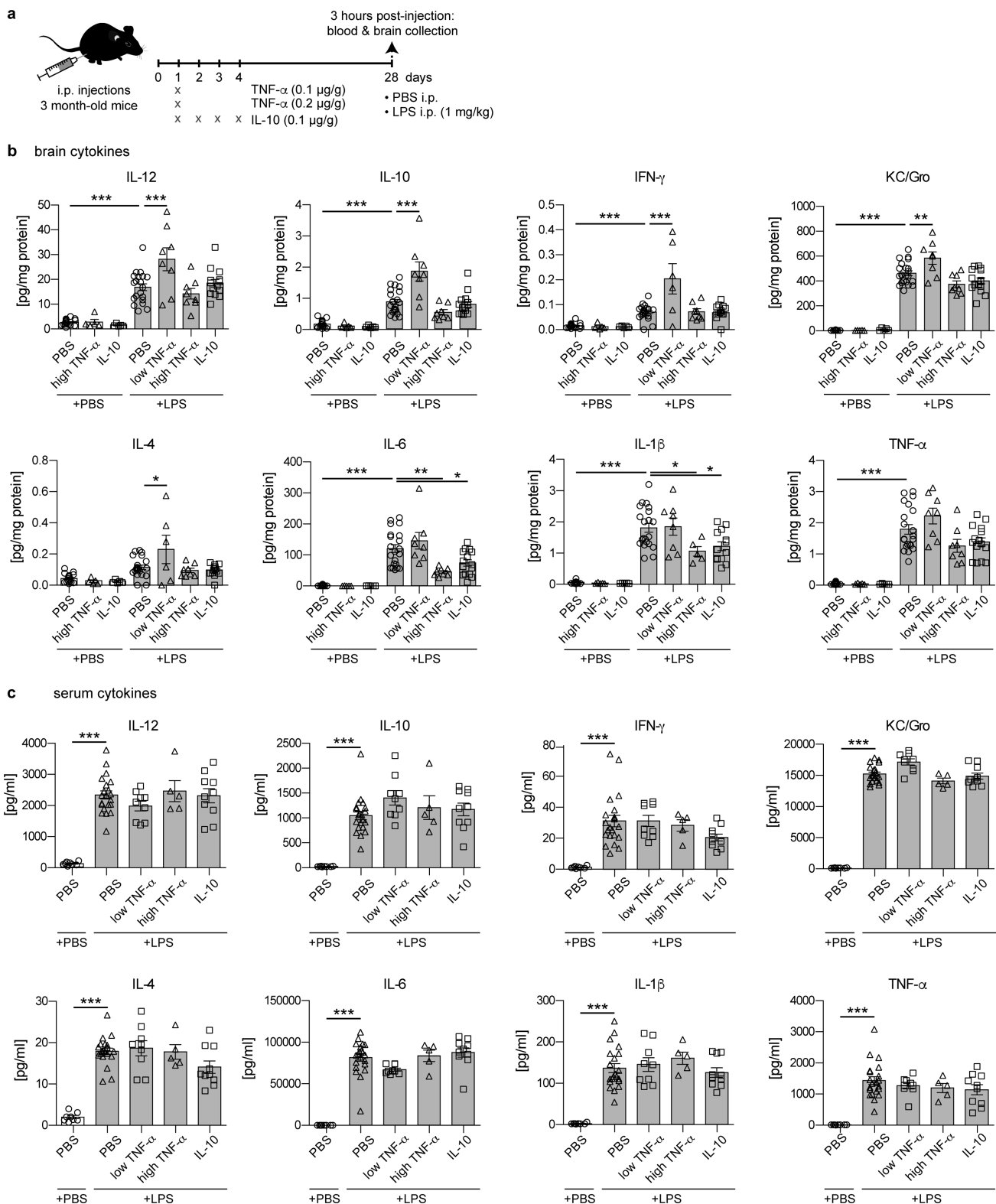
Extended Data Fig. 7 | Analysis of microglial enhancers. Microglial enhancers were analysed in 9-month-old wild-type and APP23 (APP) mice treated intraperitoneally with 1 × LPS or 4 × LPS at 3 months of age. **a**, Exemplary UCSC browser images of genomic region around the *Hif1a* gene (normalized to input and library dimension). **b**, Numbers of regions with differentially regulated H3K4me1 levels. **c**, Heatmaps of H3K4me1

regions (centred on H3K27ac peaks). **d**, Pairwise correlations between the two replicates of H3K4me1 read densities in differentially regulated regions. **e–g**, Analyses of H3K27ac levels analogous to **b–d** for H3K4me1. $n = 2$ replicates (8–10 mice per replicate); differential enhancers showed a cumulative Poisson $P < 0.0001$.

| a | motifs enriched in all active enhancers (random background): | | Known motifs | | | | De novo motifs | | | | | | | |
|---------|--|--------------|---|---------|------------------------|------------|------------------------|------------|---|------------|------------------------|---------|---------|------------|
| | | | % Sequences with Motif | | | | % Sequences with Motif | | | | | | | |
| | | | Best Match | P-value | Targets | Background | Best Match | P-value | Targets | Background | | | | |
| | | | PU.1(ETS) | 1e-624 | 26.58% | 12.58% | Sfp1 | 1e-946 | 39.65% | 19.51% | | | | |
| | | ELF5(ETS) | 1e-549 | 32.44% | 17.80% | PU.1:IRF8 | 1e-101 | 48.30% | 40.82% | | | | | |
| | | SpiB(ETS) | 1e-511 | 14.65% | 5.42% | Mef2a | 1e-91 | 10.79% | 6.90% | | | | | |
| | | ETS1(ETS) | 1e-492 | 39.47% | 24.34% | RUNX2 | 1e-78 | 41.54% | 35.16% | | | | | |
| | | ETV1(ETS) | 1e-471 | 47.55% | 31.83% | Foxo1 | 1e-70 | 29.13% | 23.68% | | | | | |
| b | comparison condition 1 2 | | known and de novo motifs enriched in enhancers of condition 1 | | | | | | known and de novo motifs enriched in enhancers of condition 2 | | | | | |
| | | | Motif | | % Sequences with Motif | | | | Motif | | % Sequences with Motif | | | |
| | | | | | Best match | P-value | Targets | Background | | | Best match | P-value | Targets | Background |
| | | | | | | | | | | | | | | |
| APP PBS | wildtype PBS | AGATCAAAAG | TCF3 | 1e-06 | 6.80% | 5.86% | TCGCAAGT | TGIF2 | 1e-14 | 69.21% | 65.97% | | | |
| | | AGCTTTGAAC | TCF3 | 1e-36 | 0.58% | 0.10% | ATGCAAGT | TGIF1 | 1e-08 | 66.14% | 63.66% | | | |
| | | ATCACGCCAT | SREBP1a | 1e-03 | 4.69% | 4.16% | AACCGCTCTCA | TGIF2 | 1e-40 | 0.30% | 0.01% | | | |
| | | CACACGGAGTGA | SREBP1a | 1e-34 | 0.33% | 0.04% | GTTCGGAGAT | IRF4 | 1e-50 | 0.50% | 0.04% | | | |
| | | AGTGAAGCA | IRF4 | 1e-03 | 11.90% | 10.96% | ATGAAGCTGC | IRF4 | 1e-23 | 0.26% | 0.03% | | | |
| | | CAAAAATTTC | IRF4 | 1e-25 | 0.27% | 0.04% | TCTGTCTCTT | Smad2/3 | 1e-29 | 1.03% | 0.32% | | | |
| | | TCATAAATAAG | Mef2c | 1e-03 | 12.17% | 11.31% | TCAGACATCSA | Smad3 | 1e-26 | 0.28% | 0.03% | | | |
| | | TGTATGTATG | Mef2b | 1e-57 | 0.53% | 0.05% | CGAGACAA | Smad2 | 1e-19 | 1.77% | 0.92% | | | |
| | | | | | | | CTATCCGTCCTT | ATF1 | 1e-26 | 0.28% | 0.03% | | | |
| | | | | | | | AGTGAAGTAT | ATF1 | 1e-25 | 0.27% | 0.03% | | | |
| | | ATTGCCAAG | CEBPb | 1e-08 | 13.68% | 12.23% | ASAGGAAGTG | PU.1(ETS) | 1e-10 | 24.57% | 22.45% | | | |
| | | ATGTGTCAA | CEBPb | 1e-02 | 16.62% | 15.82% | GGAAATGAAAGT | PU.1:IRF8 | 1e-04 | 8.96% | 8.11% | | | |
| | | TCCTTATCTG | GATA2 | 1e-06 | 15.98% | 14.62% | SGGAAATGAAAG | PU.1-IRF | 1e-03 | 38.94% | 37.61% | | | |
| | | AGATAAAGG | GATA1 | 1e-05 | 14.26% | 13.10% | CTCTCTGCTCA | PU.1(ETS) | 1e-17 | 0.37% | 0.10% | | | |
| | | AGATAAAG | GATA4 | 1e-04 | 24.00% | 22.65% | GCTGASTCAGCA | MafK | 1e-04 | 7.18% | 6.42% | | | |
| | | AGATAAAG | GATA3 | 1e-04 | 34.99% | 33.52% | AGATTCAGCAATTT | MafF | 1e-02 | 6.29% | 5.77% | | | |
| | | GCCTTATGAGCA | GATA4 | 1e-43 | 0.32% | 0.03% | TCCTGACTCA | MafA | 1e-02 | 22.71% | 21.88% | | | |
| | | AGACGTGC | HIF-1b | 1e-05 | 22.67% | 21.23% | TTGTAAGAAA | MafB | 1e-52 | 0.30% | 0.02% | | | |
| | | IACGTGCG | HIF-1a | 1e-02 | 4.20% | 3.77% | GTGTGTGACCAT | MafA | 1e-31 | 0.26% | 0.03% | | | |
| | | GTACGTAC | HIF-1a | 1e-24 | 0.69% | 0.23% | AGTGAAGCA | IRF4 | 1e-03 | 11.89% | 11.04% | | | |
| | | AGCTGTA | ARNT::HIF1A | 1e-17 | 0.27% | 0.06% | CGAGAACACCT | IRF4 | 1e-45 | 0.27% | 0.02% | | | |
| | | CCCTAGGCGAT | AP-2gamma | 1e-04 | 22.16% | 20.85% | TCGCAAGT | TGIF2 | 1e-02 | 67.78% | 66.88% | | | |
| | | ATCCCTAGGCG | AP-2alpha | 1e-03 | 16.77% | 15.77% | ATGCAAGT | TGIF1 | 1e-02 | 64.71% | 63.81% | | | |
| | | GCAGTTA | MYB | 1e-04 | 39.46% | 38.06% | AGCGGCAAG | E2F | 1e-02 | 6.96% | 6.39% | | | |
| | | ATAGTAACG | MYB | 1e-53 | 0.56% | 0.07% | ATGGCGGCA | E2F | 1e-02 | 1.93% | 1.65% | | | |
| | | AGATCCAGGTGG | MYB | 1e-53 | 0.30% | 0.01% | ATGAGTCAATG | c-Jun | 1e-02 | 6.18% | 5.71% | | | |
| | | AGGGAATG | TEAD4 | 1e-04 | 18.79% | 17.65% | ATGAGTCAATG | JunD | 1e-02 | 1.85% | 1.59% | | | |
| | | CGGCAATTC | TEAD4 | 1e-58 | 0.40% | 0.03% | | | | | | | | |
| | | ACATCAAGG | TCFL2 | 1e-03 | 2.39% | 1.98% | | | | | | | | |
| | | GACTAAGAG | TCFL2 | 1e-41 | 0.44% | 0.06% | | | | | | | | |
| | | AGGTGCTG | Smad4 | 1e-02 | 42.15% | 41.21% | | | | | | | | |
| | | TAGCGTCTG | Smad4 | 1e-36 | 0.54% | 0.10% | | | | | | | | |
| | | ATACGTG | HIF-1b | 1e-14 | 22.67% | 20.26% | ATGATGCAAT | Atf4 | 1e-04 | 5.60% | 4.89% | | | |
| | | IACGTGCG | HIF-1a | 1e-03 | 4.20% | 3.67% | GATTCGTCAG | Atf1 | 1e-47 | 0.43% | 0.04% | | | |
| | | TCACGTAA | HIF-1a | 1e-18 | 0.28% | 0.07% | GCTGATCAAGCA | MafK | 1e-04 | 7.55% | 6.76% | | | |
| | | AGGTCAAGGCA | RAR:RXR | 1e-05 | 5.71% | 4.97% | TCCTGACTCA | MafA | 1e-03 | 23.32% | 22.13% | | | |
| | | AGGTCAAGGCA | RARg | 1e-04 | 3.02% | 2.51% | AGATTCAGCAATTT | MafF | 1e-02 | 6.49% | 5.92% | | | |
| | | AGATAAAG | GATA4 | 1e-04 | 24.00% | 22.76% | ASAGGAAGTG | PU.1(ETS) | 1e-02 | 23.81% | 22.80% | | | |
| | | TCCTTATCTG | GATA2 | 1e-03 | 15.98% | 15.04% | GGAAATGAAAGT | PU.1:IRF8 | 1e-02 | 8.85% | 8.27% | | | |
| | | AGATAAAG | GATA3 | 1e-03 | 34.99% | 33.77% | SGGAAATGAAAG | PU.1-IRF | 1e-02 | 39.43% | 38.50% | | | |
| | | AGATAAAG | GATA1 | 1e-02 | 14.26% | 13.45% | AGATTCAGCAATTT | GATA3 | 1e-02 | 2.99% | 2.67% | | | |
| | | ATGTATGATGTC | MYC | 1e-59 | 0.35% | 0.01% | CAATAGCTAAG | GATA1 | 1e-51 | 0.40% | 0.02% | | | |
| | | CCATGTG | MYC | 1e-33 | 1.28% | 0.50% | TATCTCAA | GATA3 | 1e-17 | 1.95% | 1.12% | | | |
| | | GCTCATGGG | MYC | 1e-29 | 0.27% | 0.04% | | | | | | | | |
| | | AAAGCAAGATG | SpiB | 1e-15 | 13.15% | 11.19% | CCCTAGGCGAT | AP-2gamma | 1e-06 | 22.50% | 20.82% | | | |
| | | CCCTCTCTCTT | SpiB | 1e-29 | 0.31% | 0.04% | ATCCCTAGGCG | AP-2alpha | 1e-05 | 17.06% | 15.76% | | | |
| | | ATGAGTCAATG | ATF3 | 1e-04 | 15.98% | 14.83% | ATGATGCAAT | Atf4 | 1e-05 | 5.60% | 4.80% | | | |
| | | ATGAGTCAATG | ATF7 | 1e-02 | 9.72% | 9.05% | GACCGAAT | Atf4 | 1e-17 | 0.34% | 0.08% | | | |
| | | ATGAGTCAATG | ATF1 | 1e-02 | 13.83% | 13.05% | TCATAAATAG | Mef2c | 1e-05 | 12.17% | 11.05% | | | |
| | | ATGAGTCAATG | JunD | 1e-04 | 1.85% | 1.47% | CTCAAAATAG | Mef2a | 1e-02 | 10.90% | 10.14% | | | |
| | | ATGAGTCAATG | c-Jun | 1e-02 | 6.39% | 5.92% | ASATCAAAAGG | Tcf4 | 1e-04 | 11.57% | 10.53% | | | |
| | | | | | | | TCCAGGAGSCA | Tcf4 | 1e-61 | 0.45% | 0.03% | | | |
| | | | | | | | AGCTTTGAAC | Tcf3 | 1e-37 | 0.76% | 0.17% | | | |
| | | AGCTGCAAG | MEIS1 | 1e-02 | 42.34% | 41.46% | ATTGCAAG | CEBPb | 1e-03 | 13.80% | 12.81% | | | |
| | | TSACAGCAACC | MEIS1 | 1e-43 | 0.34% | 0.04% | ATAGGSCA | CEBPg | 1e-13 | 0.16% | 0.02% | | | |
| | | | | | | | GCAGTTA | Myb | 1e-02 | 39.82% | 38.67% | | | |
| | | | | | | | GCAGTCAAGAG | Myb | 1e-68 | 0.39% | 0.01% | | | |
| | | | | | | | ATTGCTATCTG | Gata5 | 1e-97 | 0.51% | 0.02% | | | |
| | | | | | | | CTTATCCTGATA | Gata6 | 1e-45 | 0.36% | 0.02% | | | |
| | | | | | | | CATAGTCT | Smad3 | 1e-27 | 6.22% | 4.28% | | | |
| | | | | | CTCTCAG | Smad2 | 1e-16 | 2.08% | 1.25% | | | | | |

Extended Data Fig. 8 | Transcription factor motif analysis of active enhancer regions. Motif analysis was performed for selected conditions to identify transcription factors involved in the differential activation of enhancers (using putative enhancer regions present in both replicates within 500 bp around enhancer peaks). **a**, For all active enhancers, motif analysis was performed using the union H3K27ac peak file and standard background (random genomic sequence). **b**, Pairwise comparisons

between conditions, using the first condition's H3K27ac peak file as input and the second condition's peak file as background. As motif enrichment was often relatively low, the analysis was focused on transcription factor (families), whose motifs occurred at least twice in 'known' (black) and 'de novo' motifs (blue). Motifs are identified by HOMER software using hypergeometric testing (no adjustment for multiple comparisons was made).



Extended Data Fig. 9 | Peripherally applied cytokines induce immune memory in the brain. **a**, Experimental design. **b**, Cytokine responses in the brain, four weeks after peripheral cytokine application ($n = 17, 5, 5, 21, 8, 8, 15$ mice from left to right). Note that TNF dose-dependently enhances (low dose) or decreases (high dose) certain cytokines. Similar to high dose

TNF, certain cytokines are also reduced by peripheral application of IL-10 four weeks earlier. **c**, Cytokine responses in the periphery are unaffected ($n = 8, 21, 9, 5, 10$ mice). Data are means \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for one-way ANOVA with Tukey correction.

Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates

Jackson Peter^{1,6}, Matteo De Chiara^{2,6}, Anne Friedrich¹, Jia-Xing Yue², David Pflieger¹, Anders Bergström², Anastasie Sigwalt¹, Benjamin Barre², Kelle Freil¹, Agnès Llored², Corinne Cruaud³, Karine Labadie³, Jean-Marc Aury³, Benjamin Istace³, Kevin Lebrigand⁴, Pascal Barbry⁴, Stefan Engelen³, Arnaud Lemainque³, Patrick Wincker^{3,5,7}, Gianni Liti^{2,7*} & Joseph Schacherer^{1,7*}

Large-scale population genomic surveys are essential to explore the phenotypic diversity of natural populations. Here we report the whole-genome sequencing and phenotyping of 1,011 *Saccharomyces cerevisiae* isolates, which together provide an accurate evolutionary picture of the genomic variants that shape the species-wide phenotypic landscape of this yeast. Genomic analyses support a single ‘out-of-China’ origin for this species, followed by several independent domestication events. Although domesticated isolates exhibit high variation in ploidy, aneuploidy and genome content, genome evolution in wild isolates is mainly driven by the accumulation of single nucleotide polymorphisms. A common feature is the extensive loss of heterozygosity, which represents an essential source of inter-individual variation in this mainly asexual species. Most of the single nucleotide polymorphisms, including experimentally identified functional polymorphisms, are present at very low frequencies. The largest numbers of variants identified by genome-wide association are copy-number changes, which have a greater phenotypic effect than do single nucleotide polymorphisms. This resource will guide future population genomics and genotype-phenotype studies in this classic model system.

The budding yeast *S. cerevisiae* is a powerful model system for understanding eukaryotic biology at the cellular, molecular and genomic levels^{1,2}. *S. cerevisiae* has recently emerged as a model in population genomics^{3–5}, because it can be found worldwide in a broad array of human-associated (for example, wine, sake, beer and other fermented beverages) and wild (for example, plant, soil and insect) biotopes. Recent years have seen a spike in the number of published *S. cerevisiae* genome sequences, which together have revealed a high level of genetic diversity and a complex population structure in this yeast^{6–12}. However, the number of available sequenced genomes from natural isolates remains limited and stands in contrast to the wealth of data on *Arabidopsis thaliana*¹³ and humans^{14,15}; this small sample size for yeast genomes has not fully captured the global evolutionary processes relevant to the species. Here we apply deep coverage genome sequencing to more than 1,000 natural *S. cerevisiae* isolates and explore their phenotypic landscape. To our knowledge, our large-scale genome analysis provides the first comprehensive view of genome evolution at different levels (for example, accounting for differences among ploidy, aneuploidy, genetic variants, hybridization and introgressions), which is challenging to obtain at this scale and accuracy for other model organisms. This sequencing effort substantiates previous hypotheses but also reveals novel aspects of *S. cerevisiae* evolutionary history. Natural *S. cerevisiae* isolates have previously proven to be a powerful tool for investigating the genotype-phenotype relationship via linkage mapping¹⁶. Although genome-wide association studies (GWAS) have led to the identification of common alleles with strong effects in other organisms, the small sample size of yeast genomes has so far prevented similar attempts for yeast. Our dataset enables GWAS and exhaustively captures genetic variants (single nucleotide polymorphisms (SNPs), copy-number variants (CNVs) and genome content), providing insights into the genetic architecture of traits and the source of missing heritability.

Species-wide genetic and phenotypic diversity

We assembled a collection of 1,011 *S. cerevisiae* isolates that maximized the breadth of their ecological and geographical origins (Supplementary Fig. 1a, b and Supplementary Table 1). We deeply sequenced 918 isolates using the Illumina paired-end strategy with a 232-fold mean coverage, and we also included 93 strains that had previously been sequenced^{6–8} (Supplementary Fig. 1c). The reads associated with each sample were mapped to the S288C reference genome and de novo assembled. A total of 1,625,809 high-quality reference-based SNPs were detected across the 1,011 genomes. Most of these SNPs are present at very low frequencies, with 31.3% of the polymorphic positions being singletons and 93% with a minor allele frequency (MAF) < 0.1 (Supplementary Fig. 2). This bias might in part be driven by the sampling scheme. Deleterious mutations as predicted by SIFT¹⁷ show the strongest bias toward rare alleles, consistent with the notion that selection prevents such SNPs from spreading in the population (Supplementary Fig. 2). In addition, we detected 125,701 small-scale insertions and deletions (indels) (up to 50 bp) with the majority exhibiting a low frequency (Supplementary Fig. 3a). Most indels are short (42.6% are 1 bp in length) and those present in coding regions are strongly biased to lengths that are multiples of three, which reflects the influence of purifying selection (Supplementary Fig. 3b, c). We also characterized CNVs by measuring the coverage ratio of each individual pangenomic open reading frame (ORF) (see below) normalized to the genome of each respective strain (Methods). CNVs are heavily enriched in subtelomeric regions, whereas internal chromosomal regions are largely copy-number stable (Supplementary Fig. 4a). Nearly all ORFs have at least one strain with a CNV. Most CNVs associated with individual ORFs are rare in the population (Supplementary Fig. 4b). Variants with high copy numbers affect only a small fraction of ORFs (Supplementary Fig. 4c) and extreme cases (> 20 copies) include

¹Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France. ²Université Côte d’Azur, CNRS, INSERM, IRCAN, Nice, France. ³Commissariat à l’Energie Atomique (CEA), Genoscope, Institut de Biologie François-Jacob, Evry, France. ⁴Université Côte d’Azur, CNRS, IPMC, Sophia Antipolis, Valbonne, France. ⁵CNRS UMR 8030, Université d’Evry Val d’Essonne, Evry, France.

⁶These authors contributed equally: Jackson Peter, Matteo De Chiara. ⁷These authors jointly supervised this work: Patrick Wincker, Gianni Liti, Joseph Schacherer. *e-mail: gianni.liti@unice.fr; schacherer@unistra.fr

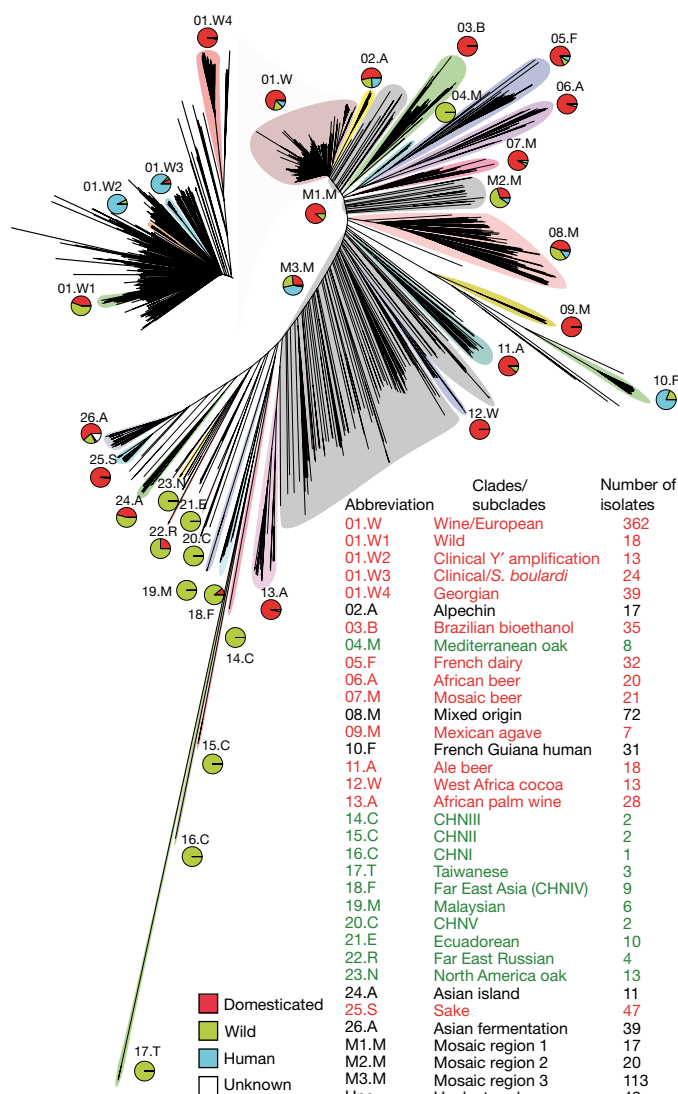


Fig. 1 | Neighbour-joining tree built using the biallelic SNPs. We identified 26 clades (numbered clockwise from 1 to 26) and three mosaic groups (M1–M3). The pie charts represent the ecological origins of the clade: domesticated (red), wild (green) and human (cyan). The colour of the clade name indicates its assignment: domesticated (red) and wild (green). The top left inset represents a magnification of the wine/European clade with four major subclades highlighted.

2 μ plasmid ORFs, mitochondrial genome, ribosomal DNA and repetitive elements such as Ty and Y'.

In parallel, 971 strains were phenotyped in different conditions that affect various physiological and cellular responses (Methods and Supplementary Table 2). In total, we analysed 34,956 phenotypic measurements that covered 36 traits, providing a comprehensive analysis of their inheritance patterns. Most of the traits vary continuously across the population in a manner consistent with their genetic complexity (Supplementary Fig. 5). However, some traits—such as resistance to copper sulfate (CuSO₄) or anisomycin¹⁸—follow a bimodal distribution model, and therefore a Mendelian inheritance pattern. Estimates of the narrow-sense heritability, h^2 , from genome-wide SNPs genotyped¹⁹ show a substantial amount of variance explained across all the traits, with a mean of 0.69, which suggests the feasibility of performing GWAS (Supplementary Fig. 6).

Population structure supports out-of-China origin

The phylogenetic tree of the 1,011 strains shows well-defined clades, loose clusters and isolated branches (Fig. 1). Most of the strains (813 in total) fall into 26 clades, and another 150 strains belong to

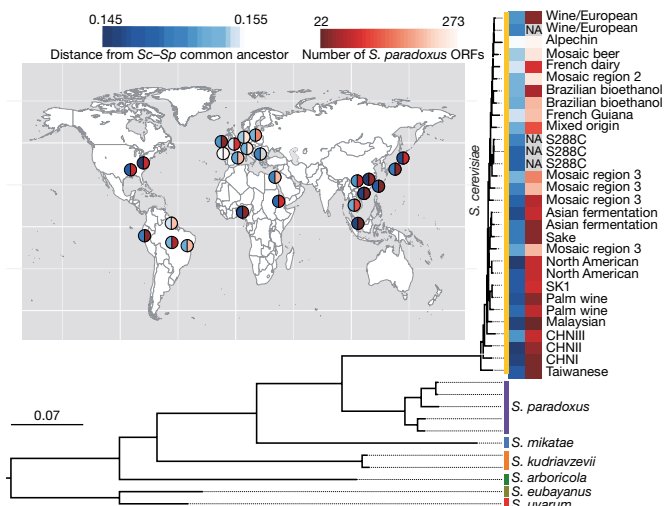


Fig. 2 | Chinese origin of *S. cerevisiae*. Maximum-likelihood rooted tree of the *Saccharomyces* complex, based on the alignment of 2,018 concatenated conserved genes. Heat maps display the distance from the last common ancestor of *S. cerevisiae* (Sc)–*S. paradoxus* (Sp) (white–blue), and the number of introgressed *S. paradoxus* ORFs (white–red). The map shows the geographical origins of the strains.

three groups of poorly related strains (Supplementary Information note 1). Our data revealed a complex pattern of genetic differentiation with distinct lineages that correlate with geography, environmental niche and the degree of human association, as has previously been reported^{8,9,11,12,20}. Domesticated and wild clades largely fall into two well-delineated sides of the tree, and are separated by a large group of mosaic strains. The main exceptions are the wild Mediterranean oak strains, which group with the domesticated clades, and the sake strains, which group with wild clades. However, the Mediterranean oak lineage groups together with the other wild lineages on the basis of ORF-content strain clustering (Supplementary Fig. 7).

We used ADMIXTURE²¹ to investigate ancestry in the genomes of individual strains. Mosaic strains are characterized by admixture from two or more lineages derived by outbreeding^{3,4} and frequently manifest as isolated branches in the phylogenetic tree. We identified three groups of mosaic strains that are mostly associated with human-related environments (Fig. 1). Population structure analysis revealed different sources of ancestry and degrees of mosaicism, consistent with multiple hybridization events (Supplementary Fig. 8). These findings underscore the role of human-driven admixture in shaping the population structure of *S. cerevisiae*.

The recent discovery of highly diverged wild Chinese lineages suggests that East Asia may represent the geographic origin of *S. cerevisiae*²². The Taiwanese wild lineage represents the most divergent population that has yet been described (average of 1.1% sequence divergence to non-Taiwanese strains). This lineage also contains an extremely divergent 2 μ plasmid that shares only 80% of identity with known plasmid variants (Supplementary Fig. 9 and Supplementary Information note 2). We used a subset of highly contiguous de novo assemblies that sample the main *S. cerevisiae* lineages and closely related *Saccharomyces* species^{23,24} to generate a rooted phylogenetic tree (Fig. 2). The outgroup species branched off near the Taiwanese and Chinese lineages, which strongly supports a Chinese origin for *S. cerevisiae*. This scenario is also consistent with the isolation of closely related *Saccharomyces* species such as *S. mikatae* and *S. arboricola*^{25,26}, which are restricted to East Asia, and the broad genetic diversity of the Japanese *S. kudriavzevii* populations²⁷. Together, these observations suggest an Asian origin for the whole *Saccharomyces* species complex. We then tested the number of out-of-China events by investigating the relationship of non-Chinese strains to the genetic structure of Chinese strains. We performed a principal component analysis on SNPs that

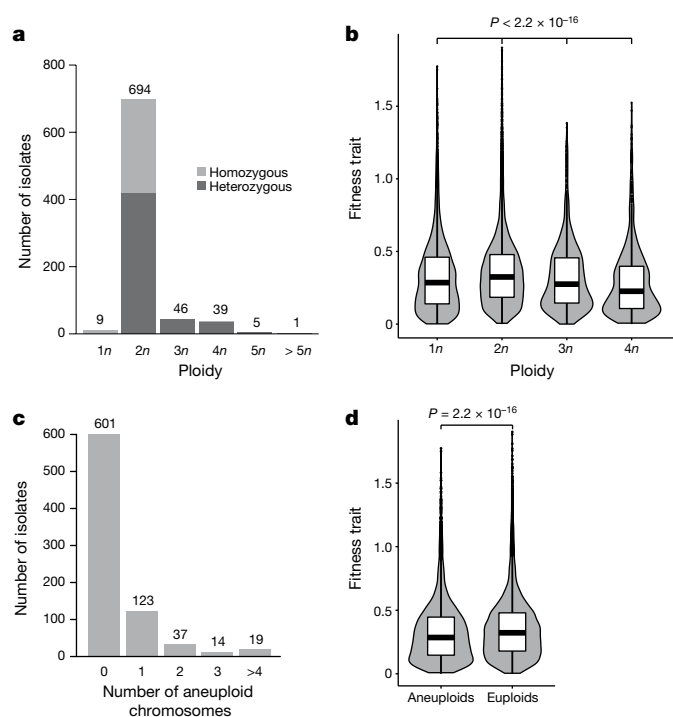


Fig. 3 | Ploidy and aneuploidy natural variation. **a**, Distribution of ploidy and fraction of heterozygous isolates. **b**, Violin plot of growth fitness trait by ploidy. Diploid isolates are globally fitter than individuals with other ploidy levels. Number of trait values for 1n isolates = 4,585; for 2n isolates = 26,249; for 3n isolates = 1,610; and for 4n isolates = 1,330. **c**, Distribution of aneuploid chromosomes per individual. **d**, Violin plot of growth fitness trait of aneuploid ($n = 6,510$) and euploid ($n = 20,719$) isolates shows a significant difference in fitness trait between the two categories. All P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. Centre lines, median; boxes, interquartile range (IQR); whiskers, $1.5 \times$ IQR. Data points beyond the whiskers are outliers.

included only the CHN I–V (Chinese isolates) and Taiwanese strains and then projected the rest of the strains onto the principal component space defined by these highly diverged lineages (Supplementary Fig. 10). Principal component 1 defines the separation of the Taiwanese strains from all other strains, consistent with the deep genetic differentiation of this lineage. Principal components 2–4 then define differentiation between different Chinese lineages. Notably, the non-Chinese strains all project onto the same part of the space, which implies that they are essentially all equally related to the different Chinese lineages. In other words, it appears that different non-Chinese strains do not derive from different Chinese lineages but instead share a single out-of-China origin.

Ploidy and aneuploidy variation by ecological origin

Variation in ploidy and aneuploidy is not uncommon in yeast, and this genomic plasticity is often described as a strategy for rapid adaptation to environmental changes^{28–30}. As 217 strains were genetically manipulated and no longer in their natural ploidy states, we assigned a relative ploidy state to the other 794 isolates (Methods), and found 9 haploids, 694 diploids and 91 isolates with higher ploidy (Fig. 3a). Our results reveal that most (about 87%) of the natural isolates are in a diploid state. Polyploid isolates (3–5n) are not frequent (approximately 11.5%) and enriched in specific subpopulations such as the beer, mixed-origin and African palm wine clades, which strongly suggests that some human-related environments have had an effect on the ploidy level (Supplementary Fig. 11). By testing the effect of ploidy on the growth fitness trait across the species, a general advantage was observed for diploid isolates compared to other types of ploidy (Fig. 3b). This result supports a general mitotic growth advantage for diploidy in yeast^{30,31} that holds across every condition tested, and shows no major environment-specific effects (Supplementary Fig. 12).

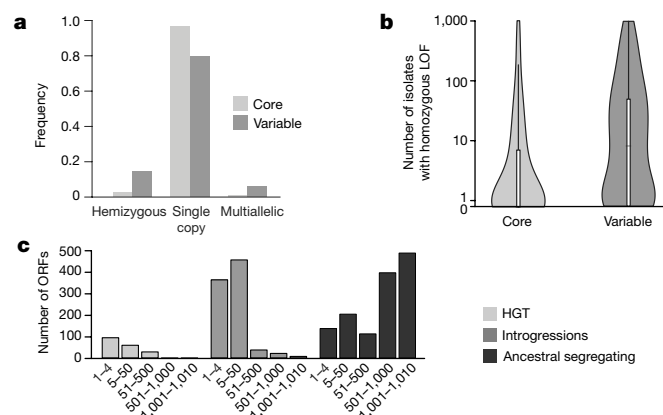


Fig. 4 | The *S. cerevisiae* pangenome. **a**, Copy number distribution for core and variable ORFs. Variable ORFs have a greater frequency of both hemizygous and multi-allelic genes. **b**, Logarithmic-scale distribution of isolates carrying loss-of-function mutations for core ($n = 4,931$) and variable ORFs ($n = 1,111$). The core genome is characterized by far fewer loss-of-function mutations compared to variable ORFs (P value = 6.45×10^{-78} , two-sided Mann–Whitney–Wilcoxon test). Centre lines, median; boxes, IQR; whiskers, $1.5 \times$ IQR. Data points beyond the whiskers are outliers. **c**, Different types of variable ORFs have marked differences in distribution.

By combining coverage analysis and allele frequency distributions, we determined the copy number of each chromosome together with instances of segmental duplications (Fig. 3c, Supplementary Fig. 13a and Supplementary Table 1). We identified a total of 342 cases of aneuploidy that affected 193 isolates (19.1% of the 1,011 strains). The most frequently observed cases of aneuploidy involve chromosomes I, III and IX and are only weakly correlated with chromosome size (Supplementary Fig. 13b). There is a strong enrichment of aneuploid strains in the sake, ale beer and mixed-origin clades (Supplementary Fig. 14). Aneuploidy is therefore not uncommon but its relationship with fitness is paradoxical. Indeed, cases of aneuploidy arise under a variety of selective regimes but lead to a decrease of global cellular fitness^{32–34}. We found a general mitotic growth advantage in euploid versus aneuploid strains, consistent with a fitness cost for chromosomal aneuploidy (Fig. 3d).

A portrait of the *S. cerevisiae* pangenome

The 1,011 genomes provided an opportunity to determine the yeast pangenome³⁵ using de novo genome assemblies and detection of non-reference material (Methods). Containing a total of 7,796 ORFs, the pangenome (Supplementary Table 3) is composed of 4,940 core ORFs and 2,856 ORFs that are variable within the population. Most core ORFs are present as a single copy per haploid genome, whereas variable ORFs show a higher frequency of both hemizygous and multi-allelic copy numbers (Fig. 4a). Analysis of the 6,081 non-redundant ORFs present in the well-annotated S288C reference genome (4,937 core and 1,144 variable ORFs) highlighted different trends. First, the distribution of variable ORFs is biased towards subtelomeric regions (Supplementary Fig. 15a), known as hotspots of gene content variation^{7,24,36}, and shows a strong functional enrichment for cell–cell interactions, secondary metabolisms and stress responses (Supplementary Table 4). Second, the core genome is characterized by lower levels of loss-of-function mutations and has a different ratio of substitution rates (ratio of nonsynonymous to synonymous polymorphisms, d_N/d_S), which reflects differences in selective constraints (Fig. 4b and Supplementary Fig. 15b). Out of the 1,072 essential genes defined in the S288C background, 123 do not belong to the core genome, although the absence of 71 of these is complemented by closely related orthologues (Supplementary Table 5).

To trace the origins of the variable ORFs, we implemented a phylogenetic approach by inspecting the evolution of each individual

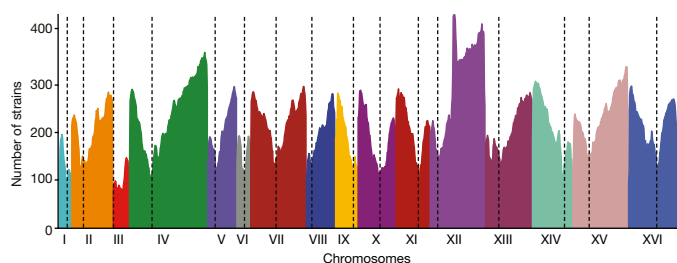


Fig. 5 | Landscape of loss-of-heterozygosity events. Density of the regions under LOH across the 16 nuclear chromosomes (I–XVI) within our population. Each colour corresponds to a chromosome, and centromere locations are represented by dotted lines.

ORF (Methods and Supplementary Fig. 16). We defined 1,380 ancestral segregating ORFs with sequence-similarity levels consistent with genome-wide species relatedness. We identified 913 introgressed ORFs, with a clear majority ($n = 885$) being unambiguously traced to a *Saccharomyces paradoxus* origin. All *S. cerevisiae* isolates carry at least one *S. paradoxus* ORF (median of 26), indicating ubiquitous gene flow between these two closely related species³⁷. We also detected 14 ORFs in the Taiwanese lineage that were introgressed from an undocumented species (Supplementary Fig. 17). The amount of introgressed content is highly variable, with an enrichment in four human-associated subpopulations in which two species might coexist and which might represent interspecific hybrid zones (Supplementary Fig. 18 and Supplementary Information note 3). By contrast, introgressions are rare in the highly diverged lineages, consistent with secondary contacts with *S. paradoxus* occurring mainly after the out-of-China dispersal (Fig. 2). There is a notable match between the geographic origins of the *S. cerevisiae* clades and the ancestry of introgressed ORFs from *S. paradoxus* (Supplementary Fig. 19 and Supplementary Information note 3).

Finally, 183 ORFs are likely to be the result of horizontal gene transfer (HGT) events from highly divergent yeast species. Introgressed ORFs tend to replace *S. cerevisiae* orthologues, which suggests that they are integrated by homologous recombination; by contrast, HGT segments localize mainly at subtelomeres. We traced the donor for 85 HGT ORFs and found an enrichment for *Torulaspora* and *Zygosaccharomyces* species. These species coexist with *S. cerevisiae* in fermentative environments, which might promote HGT events or favour their retention. We identified 6 large (38–165 kb) HGT events with most isolates retaining only small segments in complex patterns, consistent with multiple independent rearrangements leading to partial deletions of the large ancestral HGT (Supplementary Information note 3, Supplementary Figs. 20–22 and Supplementary Table 6). Together, these introgression and HGT events with distinct population frequency distributions (Fig. 4c) correspond to important evolutionary processes that have shaped the *S. cerevisiae* species genome.

Patterns of extensive loss-of-heterozygosity

S. cerevisiae is highly inbred and characterized by rare sexual cycles^{38–40}. The frequency of outcrossing has considerable effects on genome variation, and particularly on patterns of heterozygosity. Among the 794 natural isolates, 505 isolates (about 63%) are heterozygous (Fig. 3a), with the proportion varying across subpopulations and with marked differences between domesticated and wild clades (Supplementary Fig. 23), as has previously been observed⁴¹. Heterozygous sites were spread across the genomes, but we also detected large regions of loss-of-heterozygosity (LOH) and generated an accurate genome-wide LOH map (Fig. 5, Methods and Supplementary Fig. 24a). LOH events range from 2 to 56 regions per strain and represent up to 80% of the sake isolate genomes (Supplementary Fig. 24b). Although LOH levels are variable across subpopulations (Supplementary Fig. 25), we observed an overall high LOH level with 25 regions covering approximately 50% of the genome on average (Supplementary Fig. 24b, c and Supplementary

Table 7). However, LOH events are not evenly distributed along the genome (Fig. 5) and centromere-proximal regions exhibit low levels of recombination initiation, consistent with them experiencing few LOH events.

By masking the LOH regions we precisely determined the level of heterozygosity, which ranges from 0.63 to 6.56 heterozygous sites per kb (Supplementary Table 8). The distribution of the level of heterozygosity in the population is bimodal, reflecting the variability observed across subpopulations (Supplementary Fig. 26). These observed patterns are likely to be related to variation in the rate of outcrossing rate, as LOH and heterozygosity correlate with one another (Supplementary Fig. 27). The prevalence of LOH events agrees with the low outcrossing rate³⁸ and most LOH is thought to be the result of mitotic recombination⁴² or cells returning to mitotic growth after entering the meiotic phase^{43,44}. Overall, our data support the idea that *S. cerevisiae* undergoes clonal expansion followed by LOH-mediated diversification, enabling the expression of recessive alleles and generating novel allele combinations with potential effects on phenotypic diversity.

Genetic diversity and evolution by subpopulation

The comparison of genome content variation and levels of SNPs in domesticated and wild clades (Supplementary Fig. 28) shows higher SNP density (median 0.55% versus 0.41%) and lower genome content variation (median 115 ORFs that are not shared, versus 161 shared ORFs) in wild versus domesticated clades, respectively (Supplementary Fig. 28 and Supplementary Tables 9, 10). These findings suggest a shift in evolutionary mechanisms during the domestication process. The wild clades share similar genome content, and their evolution is mainly driven by the accumulation of SNPs. The specific artificial environments colonized by domesticated clades probably promote rapid ORF expansion and/or loss, leading to considerable variation in genome content and CNVs. Some domesticated clades also exhibit high copy numbers of Ty1 and Ty2 transposable elements (Supplementary Fig. 29).

We investigated evolutionary patterns across multiple *S. cerevisiae* subpopulations using 19 clades that contained at least 10 isolates. By determining the ploidy and measuring the genome-wide levels of genetic diversity (π and θ_w), the MAF spectrum and Tajima's D values, our results revealed distinct evolutionary histories across subpopulations (Supplementary Table 11). We also estimated the timings of the *S. cerevisiae* out-of-China and domestication events (Supplementary Information note 4).

Analyses of human-related *S. cerevisiae* subpopulations provide strong evidence for various independent and lineage-specific domestication events. Although *S. cerevisiae* beer isolates are polyphyletic, they are characterized by higher ploidy ($\geq 3n$) and a higher number of aneuploidy events than other domesticated lineages. In addition, the genetic diversity of beer isolates is very high ($\pi = 2.8 \times 10^{-3}$ on average) and the number of heterozygous sites they possess is elevated (ranging from 17,807 to 34,203 heterozygous sites on average). Finally, LOH regions represent a small proportion of the beer genomes (11% on average). The independent but convergent domestication processes undergone by beer isolates are, therefore, marked by genome-level modification and high nucleotide variation.

By contrast, wine and sake isolates are primarily diploids, monophyletic and have limited genetic diversity (π values of 1×10^{-3} and 0.8×10^{-3} , respectively). The wine cluster is characterized by a strong bias towards low-frequency polymorphisms, with more than 95% of the polymorphic sites with MAF < 0.1 . In addition, wine isolates harbour low heterozygosity and extensive LOH, which suggests a low outcrossing rate. All of these observations indicate that wine isolates experienced a population expansion after a domestication bottleneck. The effect of the domestication event on the sake genomes is very similar. Indeed, these genomes have low levels of genetic diversity, low heterozygosity and extensive LOH regions. However, the sake subpopulation does not exhibit a bias towards low-frequency alleles (Tajima's D value of 0.0481), reflecting their more recent origin (Supplementary Information note 1)⁴⁵.

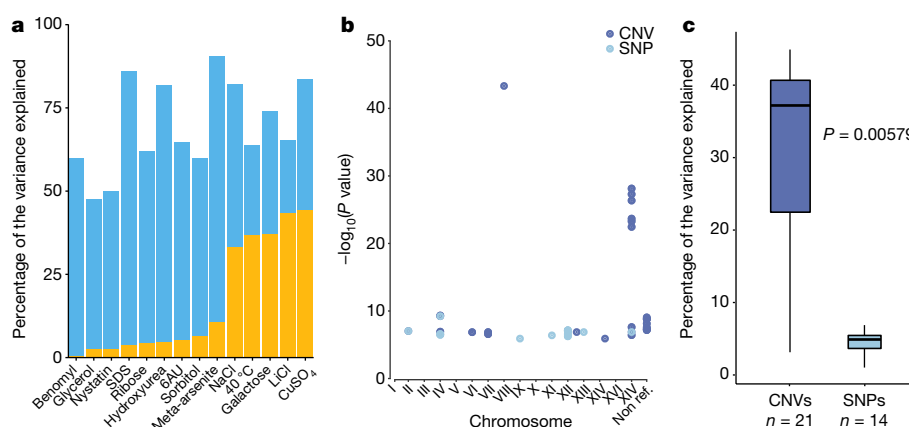


Fig. 6 | Genotype-phenotype relationship in *S. cerevisiae*. **a**, Narrow-sense heritability (blue) and phenotypic variance explained (yellow) for phenotypes with associated variants. **b**, Association scores of the detected genetic variants across the 16 chromosomes and the non-reference ORFs. **c**, Variance explained by CNVs and SNPs associated with traits.

New insights into the genotype-phenotype relationship

Natural *S. cerevisiae* isolates have been a powerful tool for linkage mapping, revealing an large number of quantitative trait loci, over 100 quantitative trait genes and 50 quantitative trait nucleotides¹⁶. We investigated the allele frequency associated with these polymorphisms, which underlie quantitative trait variation (Supplementary Table 12). Among 36 well-characterized quantitative trait nucleotides, 24 were found with at a frequency of lower than 5%, and were therefore undetectable using a GWAS strategy. This strong bias towards rare alleles for these polymorphisms is consistent with the overall MAF spectrum. Consequently, this result highlights that a fraction of the missing heritability of complex traits might be explained by rare variants in species with an allele frequency distribution similar to that of yeast.

The high genetic diversity ($\pi = 3 \times 10^{-3}$) and low linkage disequilibrium ($LD_{1/2} = 500$ bp) (Supplementary Fig. 30) among *S. cerevisiae* isolates indicates that this species could represent a powerful resource for GWAS. We built a matrix of genetic variants that included comprehensive sets of SNPs, CNVs and non-reference variable ORFs. Our matrix contains a total of 82,869 SNPs and 925 CNVs, which represents a dense map with, on average, one marker every 143 bp. We estimated genome-wide phenotype heritability¹⁹ and revealed that traits are not stratified by subpopulations (Fig. 6a, Supplementary Figs. 31, 32 and Supplementary Tables 13, 14).

We performed a mixed-model association¹⁹ and detected 35 genetic variants associated with 14 conditions (22 CNVs versus 13 SNPs), with an enrichment and high association scores for CNVs (Fig. 6b, Supplementary Fig. 33 and Supplementary Table 15). In addition, four of the variants we detected are linked to non-reference ORFs. For five traits, the phenotypic variation explained is greater than 25% (Fig. 6a). CNVs explained larger proportions of trait variance compared to SNPs (a median of 36.8% and 4.49% of the variance explained, respectively; Fig. 6c). As an example, we found the *CUP1* gene strongly associated with resistance to copper sulfate (P value = 4.85×10^{-44}) (Supplementary Fig. 33 and Supplementary Table 15). Amplification of this locus strongly contributes to the resistance to high concentrations of copper and cadmium⁴⁶ with copy number variation alone explaining 44.5% of phenotypic variation. Our GWAS analyses, which included an exhaustive catalogue of genome content and CNVs, highlighted the overall importance of these genetic variants for phenotypic diversity. The high number of associated CNVs is consistent with the notion that these variants can contribute to a large amount of phenotypic variation⁴⁷.

Conclusion

Our whole-genome sequencing of 1,011 isolates, combined with our phenotyping efforts, provides a detailed view of *S. cerevisiae* variation.

Association scores and variance explained are higher for CNVs compared to SNPs (P value = 0.00579, two-sided Mann-Whitney-Wilcoxon test). Centre lines, median; boxes, IQR; whiskers, $1.5 \times$ IQR. Data points beyond the whiskers are outliers.

This resource has revealed previously undescribed evolutionary history as well as the driving forces of genome evolution, and has provided insights into the genotype-phenotype relationship. Our study lays the foundation for GWAS in *S. cerevisiae* and provides a population genomic resource at a scale that matches those of other model organisms^{13,14}. The difference between the estimated genome-wide heritability and explained phenotypic variance gives an overview of the extent of missing heritability^{48,49}. Many SNPs are present at low frequencies, which echoes observations previously made in human GWAS¹⁵ and raises the question of whether rare SNPs have an important role in modulating the phenotypic landscape. Furthermore, the comprehensive characterization of the species pangenome can further improve the genotype-to-phenotype map. The availability of end-to-end genome assemblies will enable the organization of such a dataset with genome graphs⁵⁰. This collection of genetic and phenotypic variants will therefore enable novel functional approaches in a powerful model system.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0030-5>.

Received: 4 May 2017; Accepted: 9 January 2018;

Published online 11 April 2018.

- Botstein, D. & Fink, G. R. Yeast: an experimental organism for 21st century biology. *Genetics* **189**, 695–704 (2011).
- Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–345 (2009).
- Liti, G. et al. Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
- Peter, J. & Schacherer, J. Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* **33**, 73–81 (2016).
- Skelly, D. A. et al. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* **23**, 1496–1504 (2013).
- Bergström, A. et al. A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* **31**, 872–888 (2014).
- Strope, P. K. et al. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* **25**, 762–774 (2015).
- Almeida, P. et al. A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* **24**, 5412–5427 (2015).
- Zhu, Y. O., Sherlock, G. & Petrov, D. A. Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)* **6**, 2421–2434 (2016).
- Gallone, B. et al. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**, 1397–1410.e16 (2016).

12. Gonçalves, M. et al. Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* **26**, 2750–2761 (2016).
13. The 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
14. The 1001 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
15. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
16. Fay, J. C. The molecular basis of phenotypic variation in yeast. *Curr. Opin. Genet. Dev.* **23**, 672–677 (2013).
17. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protocols* **4**, 1073–1081 (2009).
18. Hou, J. et al. The hidden complexity of Mendelian traits across natural yeast populations. *Cell Reports* **16**, 1106–1114 (2016).
19. Lippert, C. et al. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
20. Ludlow, C. L. et al. Independent origins of yeast associated with coffee and cacao fermentation. *Curr. Biol.* **26**, 965–971 (2016).
21. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
22. Wang, Q. M., Liu, W. Q., Liti, G., Wang, S. A. & Bai, F. Y. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* **21**, 5404–5417 (2012).
23. Borneman, A. R. & Pretorius, I. S. Genomic insights into the *Saccharomyces sensu stricto* complex. *Genetics* **199**, 281–291 (2015).
24. Yue, J. X. et al. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* **49**, 913–924 (2017).
25. Boynton, P. J. & Greig, D. The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* **31**, 449–462 (2014).
26. Liti, G. et al. High quality *de novo* sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics* **14**, 69 (2013).
27. Hittinger, C. T. et al. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* **464**, 54–58 (2010).
28. Hose, J. et al. Dosage compensation can buffer copy-number variation in wild yeast. *eLife* **4**, e05462 (2015).
29. Sunshine, A. B. et al. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol.* **13**, e1002155 (2015).
30. Gerstein, A. C., Chun, H. J., Grant, A. & Otto, S. P. Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* **2**, e145 (2006).
31. Zörgö, E. et al. Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* **9**, e1003388 (2013).
32. Torres, E. M. et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007).
33. Yona, A. H. et al. Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl Acad. Sci. USA* **109**, 21010–21015 (2012).
34. Tan, Z. et al. Aneuploidy underlies a multicellular phenotypic switch. *Proc. Natl Acad. Sci. USA* **110**, 12367–12372 (2013).
35. Tettelin, H. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl Acad. Sci. USA* **102**, 13950–13955 (2005).
36. Brown, C. A., Murray, A. W. & Verstrepen, K. J. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* **20**, 895–903 (2010).
37. Dunn, B., Richter, C., Kvitek, D. J., Pugh, T. & Sherlock, G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* **22**, 908–924 (2012).
38. Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* **38**, 1077–1081 (2006).
39. Tsai, I. J., Bensasson, D., Burt, A. & Koufopoulos, V. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc. Natl Acad. Sci. USA* **105**, 4957–4962 (2008).
40. Friedrich, A., Jung, P., Reisser, C., Fischer, G. & Schacherer, J. Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol. Biol. Evol.* **32**, 184–192 (2015).
41. Magwene, P. M. et al. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **108**, 1987–1992 (2011).
42. Llorente, B., Smith, C. E. & Symington, L. S. Break-induced replication: what is it and what is it for? *Cell Cycle* **7**, 859–864 (2008).
43. Laureau, R. et al. Extensive recombination of a yeast diploid hybrid through meiotic reversion. *PLoS Genet.* **12**, e1005781 (2016).
44. Brion, C. et al. Variation of the meiotic recombination landscape and properties over a broad evolutionary distance in yeasts. *PLoS Genet.* **13**, e1006917 (2017).
45. Ohnuki, S. et al. Phenotypic diagnosis of lineage and differentiation during sake yeast breeding. *G3 (Bethesda)* **7**, 2807–2820 (2017).
46. Fogel, S. & Welch, J. W. Tandem gene amplification mediates copper resistance in yeast. *Proc. Natl Acad. Sci. USA* **79**, 5342–5346 (1982).
47. Conrad, D. F. et al. Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
48. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
49. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
50. Paten, B., Novak, A. M., Eizenga, J. M. & Garrison, E. Genome graphs and the evolution of genome inference. *Genome Res.* **27**, 665–676 (2017).

Acknowledgements We thank our colleagues (listed in Supplementary Table 1) for their gifts of yeast strains, and particularly the DBVPG and NCYC yeast collections for donating large batches of strains and assistance with identification. We thank B. Dujon and D. Speed for discussions. The sequencing effort was funded by France Génomique (ANR-10-INBS-09-08). This work was also supported by the National Institutes of Health (NIH) grant R01 GM101091-01 to J.S. and by ATIP-Avenir (CNRS/INSERM), Fondation ARC (SFI20111203947), the French National Research Agency (ANR-13-BSV6-0006-01, ANR-16-CE12-0019, ANR-11-LABX-0028-01), Cancéropôle PACA (AAP emergence) and DuPont Young Professor Award to G.L. J.P. was supported by Ministère de l'Enseignement Supérieur, J.-X.Y. by Fondation ARC (PDF20150602803) and B.B. by La Ligue Contre le Cancer (GB-MA-CD-11287). We also acknowledge the IRCAN Flow Cytometry Facility CytoMed. J.S. is a Fellow of the University of Strasbourg Institute for Advanced Study (USIAS) and a member of the Institut Universitaire de France.

Reviewer information *Nature* thanks D. Gresham, G. Sherlock and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.F. and A.L.I. prepared the DNA samples. A.S. and B.B. performed the phenotyping. J.P., M.D.C., A.F., J.-X.Y., D.P., A.B. and K.L. analysed the data. C.C., K.L., J.-M.A., B.I., S.E., A.L. and P.W. sequenced the genomes and produced the data. P.B., P.W., G.L. and J.S. contributed tools and materials. G.L. and J.S. conceived and coordinated the project. G.L. and J.S. wrote the paper with the input of J.P., M.D.C. and A.F.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0030-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to G.L. or J.S. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

***S. cerevisiae* sequenced isolates.** The isolates included in this project were carefully selected to be representative of the *S. cerevisiae* whole species. All the isolate details, including ecological and geographical origins, providers and references, are listed in Supplementary Table 1. We maximized the isolate ecological origins by including both human-associated environments such as wine and sake fermentation, brewing and dairy products, as well as natural environments such as soil, insects, tree exudate and fruit. Geographical origins are also highly diverse and have a worldwide distribution (Supplementary Table 1). In addition to the 918 isolates provided by research laboratories and yeast collections, we included 93 strains sequenced in previous studies^{6–8}, to give a total of 1,011 samples analysed in this study. We sought to keep the isolates in their natural state before sequencing to provide a global picture of the ploidy and level of heterozygosity. However, among the 918 selected isolates, 124 were non-natural haploid with the *HO* gene deleted and the 93 external isolates were genetically manipulated and made homozygous before sequencing.

Sequencing and quality filtering. Yeast cell cultures were grown overnight at 30 °C in 15 ml of YPD medium to early stationary phase. Total genomic DNA was subsequently extracted using MasterPure Yeast DNA Purification Kit and Genomic Illumina HiSeq 2000 sequencing libraries were prepared for 918 strains with an insert size between 300 and 600 bp. Ten libraries were multiplexed per Illumina HiSeq2000 lane and subjected to paired-end sequencing, producing reads of 102 bases.

An in-house quality control process was applied to the reads that passed the Illumina quality filters. Illumina sequencing adapters and primers sequences were removed from the reads and the low-quality nucleotides ($Q < 20$) were discarded from both ends of the reads. Reads shorter than 30 nucleotides after trimming were removed. These trimming and removal steps were achieved using software designed in-house, based on the FastX package. The last step identifies and discards read pairs that mapped to the phage phiX genome (GenBank code: NC_001422.1) using SOAP⁵¹. A total of 3.35 Tb of high-quality genomic sequence was generated with a mean sequencing depth of $232 \times$ per isolate (ranging from $50 \times$ to $1,014 \times$, Supplementary Fig. 1c). For the publically available Illumina paired-end reads related to 93 strains (see '*S. cerevisiae* sequenced isolates'), the mean sequencing depth is $169 \times$ (from $20 \times$ to $570 \times$, Supplementary Fig. 1c).

Reads mapping and variant calling. For each isolate, the reads were mapped to the *S. cerevisiae* S288C reference genome (version R64-1-1) with Burrows–Wheeler Aligner (BWA v.0.7.4-r385)⁵², using default parameters. Duplicated reads were marked with Picard-tools (v.1.124) (<http://picard.sourceforge.net>) and local realignment around indels and variant calling were performed with GATK (v.3.3-0)⁵³. Default parameters were applied except for the realignment step (GATK IndelRealigner), in which the following parameters were set: ‘–maxReadsForConsensuses 500–maxReadsForRealignment 40000–maxConsensuses 60–maxPositionalMoveAllowed 400–entropyThreshold 0.2’. GATK Variant Annotator was run to add allele balance information in the .vcf files.

Ploidy, types of aneuploidy and segmental duplications. The natural ploidy of the 794 natural isolates (see '*S. cerevisiae* sequenced isolates'), as well as their aneuploidy and segmental duplication content were investigated by combining three complementary approaches:

First, measurement of the cell DNA content using high-throughput flow cytometry: DNA content was analysed using a propidium iodide (PI) staining assay. Cells were first pulled out from glycerol stocks in liquid YPD in 96-well plates (30 °C, overnight). Five microlitres of the culture were transferred into 195 µl of fresh YPD and incubated for 8 h at 30 °C. Then, 3 µl were taken and resuspended in 100 µl of cold 70% ethanol. Cells were fixed overnight at 4 °C, washed twice with PBS, resuspended in 100 µl of staining solution (15 µM PI, 100 µg/ml RNase A, 0.1% v/v Triton-X, in PBS) and finally incubated for 3 h at 37 °C in the dark. Ten thousand cells for each sample were analysed on a FACS-Calibur flow cytometer using the HTS module for processing 96-well plates. Cells were excited at 488 nm and fluorescence was collected with a FL2-A filter. The distributions of both FL2-A and FSC-H values have been processed to find the two main density peaks, which correspond to the two cell populations in G1 and G2. The peaks were detected using the densityClust R package after removing the cells reaching the FACS saturation (either FLS-A or FSC-H values equal to 1,000). We categorized the values of FLS-A, which correlate with the DNA quantity, to estimate the ploidy according to the following scheme: strains with G1 cell values between 39 and 181 and G2 values between 148 and 255 were labelled as haploid; strains with G1 cell values between 145 and 265 and G2 values between 295 and 500 were labelled as diploid; strains with G1 cell values between 245 and 355 and G2 values between 500 and 700 were labelled as triploid; strains with G1 cell values between 295 and 500 and G2 values between 700 and 905 were labelled as tetraploid; strains with G1 cell values between 395 and 605 and G2 values over 905 were labelled as over $4n$; strains with other combinations of values have been manually evaluated.

Second, study of sequencing coverage: systematic analysis of the coverage depth along the genome was performed with 1-kb non-overlapping sliding windows, which enabled the survey of chromosomal copy number variations as well as segmental duplications. The ratio between the coverage of the aneuploid chromosomes and the rest of the genome was also used to validate the ploidy of isolates.

Third, investigation of the allele balance ratio associated with heterozygous SNPs, as heterozygous sites should fit an expected range of ratios according to the copy number of the chromosome being considered (see 'SNPs filtering and matrix').

The precise locations of segmental duplications were manually investigated in the .vcf files (Supplementary Table 16).

SNP filtering and matrix. For each sample, variants were first called with GATK HaplotypeCaller (see 'Reads mapping and variant calling'). At this stage, isolates with less than 5% of heterozygous sites (average percentage of heterozygous sites detected in a sample of 104 haploid and/or homozygous diploid isolates) were considered as homozygous. The raw files were then post-processed to deal with highly confident variants to be included in our complete SNPs matrix, based on both coverage and allele balance information:

First, a minimal coverage depth of $50 \times$ was required for a SNP to be retained for the 918 isolates that were sequenced in this study; this coverage depth was lowered to $10 \times$ for the other 93 previously sequenced isolates.

Second, for the haploid and homozygous isolates ($< 5\%$ of heterozygous sites), the fraction of heterozygous SNPs detected was considered as representing false positives and was therefore filtered out.

Third, for heterozygous isolates, heterozygous sites were filtered according to their allele balance ratio (ABHet). The thresholds for allele balance ratios were determined according to the allelic frequency distribution all over the heterozygous samples at each level of ploidy (from $2n$ to $5n$). A heterozygous site was rejected when its ratio did not fit the expected range according to the copy number of the chromosome being considered (or region, in the case of segmental duplication).

The joint calling method of GATK was run with the cleaned .vcf files to create a complete genotyping matrix (.gvcf format, see 'Data availability'). This matrix of SNPs included 1,625,809 segregating sites across our 1,011 isolates (Supplementary Table 1).

SnEff (v.4.1)⁵⁴ was used to annotate and predict the effect of the variants. Non-synonymous SNPs predicted as deleterious by SIFT (v.5.2.2)¹⁷ as well as non-sense mutations were considered as deleterious for protein function. Insertions and deletions were considered to cause frame shifts when their sum produced a number not divisible by three in a single gene.

A sequence representative of each isolate was constructed by inserting these filtered SNPs into the reference sequence with GATK FastaAlternateReferenceMaker (see 'Data availability').

Pangenome. De novo genome assembly. We used Abyss software (v.1.5.2)⁵⁵ with the option ‘-k 64’ to produce the de novo assemblies (see 'Data availability'). The pre-assembly filtering step was performed with condetri v.2.2 to remove the 6 bases closest to the 5' end and to discard low-quality 3'-end bases of reads. The resulting assemblies had a median N50 of 136 kb and a median number of contigs of 3,259. The median length of the genome is 12.1 Mb and the median GC content is 38.06 (Supplementary Table 17).

Detection of non-reference material. We set up a custom pipeline to identify non-reference genome material. Each genome was aligned to the reference sequence (S288C, version R64-1-1) using blastn with the following settings: ‘-gapopen 5 -gapextend 5 -penalty -5 -reward 1 -evalue 10 -word_size 11 -no_greedy’.

The CDH and CFH strains were excluded from the identification of non-reference genome material owing to the presence of *Staphylococcus epidermidis* contamination. The sequences aligned with an identity greater than 95% were divided in three categories to be further processed. If the aligned sequence belonged to contigs shorter than 100 bp or if the aligned sequence was up to 200 bp and belonged to a contig with a length that was shorter than the length of the alignment plus 75 bp, the contig was discarded. If the aligned sequence was in the range 200–1,500 bp only the aligned sequence was discarded. If the aligned sequence was longer than 1,500 bp, it was divided into segments of 250 bp. Each sub-sequence was aligned again to the reference and discarded if found with an alignment identity of over 95% on an alignment length of at least 187 bp (75% of the subsequence length). After this step the relative position of the retained sequence has been evaluated. If two or more of them belonged to the same contig and were separated each other by less than 100 bp, the sequence from the starting of the first one to the end of the final one was kept as a whole. In the subsequent step, all the kept sequences from the 1,011 genomes were sorted for length in decreasing order. The set of sequence was then aligned against itself (with the same criteria as the first step) to eliminate repeated elements. When two sub-sequences were found to have an alignment identity of over 95%, the one belonging to the shorter sequence was eliminated. The process led to 12,325 sequences for 9.3 total Mb.

Annotation of non-reference material. To annotate ORFs in dispensable regions, we set up an integrative yeast gene annotation pipeline by combining different existing annotation approaches, which gave rise to an evidence-leveraged final annotation²⁴. We independently ran the three individual components (RATT⁵⁶, YGAP⁵⁷ and MAKER⁵⁸) for gene annotation, and subsequently integrated their results using EVM⁵⁹. Proteomes of the *Saccharomyces* species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. arboricolus*, *S. uvarum* and *S. eubayanus*) were retrieved and used in our annotation pipeline to provide protein alignment support for annotated gene models^{60,61}.

Pangenome definition. We compiled the pangenome by adding the 2,245 non-reference ORFs annotated here to the 6,713 genomic reference ORFs listed in the set 'orf_genomic_all' from the *Saccharomyces* Genome Database (SGD) database (updated 13 January 2015, https://downloads.yeastgenome.org/sequence/S288C_reference/orf_dna/). Three RDN genes were also added (*RDN18-1*, *RDN25-1* and *RDN37-1*) from the set 'rna_genomic' available from the SGD database (https://downloads.yeastgenome.org/sequence/S288C_reference/rna/). We applied a graph-based pipeline to this set of ORFs, to remove duplicate and closely related sequences. This step also removed overlapping ORFs present in the 'orf_genomic_all' SGD dataset. A disconnected graph was created in which each node is an ORF and each edge represents an alignment identity of over 95% on at least 75% of the sequence of the smaller ORF in the couple. Each connected subgraph represents a single ORF family. For each of these families a representative has been chosen. The connectivity has been computed for each node. The first choice for the representative was the most central, non-dubious reference ORF, if any of them were present. The second choice was the most central reference ORF; if only non-reference ORFs were present in the family, the most central of these was taken.

This led to a catalogue of 7,796 non-redundant ORFs (see 'Data availability'), which represent the *S. cerevisiae* pangenome (Supplementary Table 3). Among the reference ORFs, the number of similar ORFs collapsed for each final ORFs has a wide range (up to 67, which is the cluster of Ty1 elements), although usually this number does not exceed 2. Other large clusters are the Y' elements one (59 ORFs) and the Ty2 elements (27 ORFs). Out of the 6,713 reference ORFs, 5,679 were not redundant and 1,032 were collapsed into 402 unique ORFs; 89% of these unique ORFs ($n = 357$) are duplicated ORFs.

Pangenome CNVs. To assess the copy number of each ORF of the pangenome, we mapped the reads from each strain to the pangenomic ORFs with BWA, using default parameters and the option -U 0. The result was then filtered using samtools view with the options -bS 20 -F 260. The median coverage for each ORF was taken as coverage for the ORF in the specific isolate. The ratio between the values of individual ORFs and the values of genome coverage on the reference of the isolate (as the median of the median coverage for each nuclear chromosome) was considered as the copy number for the haploid genome (see 'Data availability').

The mapping was also used as a confirmation step for the presence of the ORFs in each strain, leading to the identification of 4,940 ORFs present in the 1,011 strains of the collection, representing the core genome plus 2,856 ORFs present in different subsets of the population. Fifty-two ORFs were removed because they were present in single strains with low coverage (~10% of the genome wide coverage) and were likely to be contaminations from *Escherichia coli* and *Clavispora lusitaniae*. Eighty-nine other ORFs that did not have sufficient coverage were kept in the pangenome, but were not used for subsequent analyses owing to poor mapping. Three of the core ORFs are present but not annotated in the S288C reference and were annotated by our annotation pipeline as 584-snap_masked-1700-AIE_1, 610-snap_masked-2999-BGP_1 and 611-snap_masked-3001-BGP_1.

To evaluate the difference between domesticated clades and wild clades, we normalized the data by calculating the clade copy-number median for each ORF to avoid sample bias. The distributions of medians in the domesticated and wild clades were then compared using the Mann-Whitney-Wilcoxon test (R function wilcox.test) (Supplementary Fig. 28 and Supplementary Tables 18, 19).

Inference of pangenome origin. We constructed a local ORFs database for 57 representative species that deeply probed both closely related *Saccharomyces* species as well as a highly divergent yeast species⁶² (Supplementary Table 20). In addition, we added the ORFs of 12 representative *S. cerevisiae* and *S. paradoxus* strains with complete genome sequenced by long reads (PacBio)²⁴. For each annotated variable ORF, we first performed a BLASTN search ('-evalue 1E-6') against this local ORFs database to find its best hit. ORFs without hits in our local yeast ORFs database underwent a further round of BLAST searching ('-evalue 1E-6') against the NCBI non-redundant database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>). Based on the sequence identity and query coverage of the top hits, we classified the variable genes into different categories.

d_N/d_S . For all isolates, sequences of the protein-coding genes were inferred from the filtered SNPs and inserted into the reference sequence with GATK FastaAlternateReferenceMaker. For each gene, the coding sequences were aligned and the ratio of nonsynonymous to synonymous polymorphisms (d_N/d_S) was

computed with the yn00 program in PAML software⁶³. Median values were used for comparison.

Genomic diversity characterization. *Genomic and genetic distances.* The 1,544,489 biallelic segregating sites were used to construct a neighbour-joining tree (Fig. 1), using the R packages ape and SNPrelate. The .gvcf matrix was first converted into a .gds file and individual dissimilarities were estimated for each pair of individuals with the snpgdsDiss function. The bionj algorithm was then run on the distance matrix that was obtained.

The genomic content distance (see 'Data availability') has been calculated as the number of ORF differences in the pangenome presence/absence profile (that is, the number of ORFs present in only one strain for each pairwise strain comparison) (see 'Data availability').

Genetic diversity. As an estimate of the scaled mutation rate, we computed π , the average pairwise nucleotide diversity θ_w , the proportion of segregating sites and Tajima's D value, which represents the difference between π and θ_w .

Variscan 2.0⁶⁴ was run (runmode = 12, 10-kb non-overlapping windows) on multiple alignments of the concatenated chromosomes that were representative of the isolates.

Model-based ancestry. Model-based ancestry estimation was performed on the biallelic SNPs using ADMIXTURE v.1.23²¹ in unsupervised mode.

Principal component analysis. Principal components analysis on the biallelic SNPs was performed using EIGENSOFT v.6.0.1. The '-w' argument was used to calculate the principal components using only a subset of the samples, with the remaining samples then being projected onto the resulting components.

Discriminant analysis of principal components. The matrix of presence/absence of ORFs in the population has been analysed using the discriminant analysis of principal components (DAPC) algorithm implemented in the R package adegenet 2.0.1. DAPC describes clusters by maximizing the between-cluster variance while minimizing the within-cluster variance. The number of components retained for the principal component analysis calculation was 150, accounting for > 88% of total variance. For the subsequent DAPC calculation, the alpha-score indicates 25 as the optimal number of discriminant principal components to be retained. Clustering was performed using the K-means with different number of groups ($n = 5, 10, 15, 20, 25, 30, 35, 40, 45$ and 50).

Linkage disequilibrium. The Plink package⁶⁵ was used to compute r^2 , the correlation coefficient between pairs of loci that stands as a measure of association for linkage disequilibrium. All pairs of polymorphic sites were investigated through .map and .ped files generated with vcftools⁶⁶, excluding SNPs with a MAF lower than 5%.

We averaged r^2 based on the SNP distance (100-bp intervals) over 25-kb regions and calculated the half-length of r^2 , which is the distance at which linkage disequilibrium decays to half of its maximum value.

Loss of heterozygosity. Heterozygous isolates were investigated for LOH regions with an R script generated in-house (see 'Data availability'). Regions over 50 kb with less than 10 heterozygous sites per 50 kb were considered to be under LOH (Supplementary Table 7).

Saccharomyces rooted tree. To construct the tree, we used 22 *S. cerevisiae* isolates representative of the species genetic diversity that were sequenced with Oxford Nanopore technology⁶⁷. We annotated these 22 assemblies with the pipeline described above. The annotated protein-coding genes were pooled together with the *S. cerevisiae* reference genome (SGD R64-1-1) and another 18 yeast strains for orthology identification. These 18 other yeast strains included 7 *S. cerevisiae* strains, 5 *S. paradoxus* strains and 6 out-group strains from other *Saccharomyces* yeast species as previously described²⁴. The orthology identification was carried out using Proteinortho (v.5.15)⁶⁸ and synteny information was considered (the PoFF feature of Proteinortho). This leads to the delineation of 2,018 1-to-1 orthologous groups across all the 41 sampled genomes. For each orthologous group, the protein sequences across the 41 strains were aligned with MUSCLE (v.3.8.1551)⁶⁹, and the resulting protein alignment was further used to guide the corresponding CDS alignment using PAL2NAL (v.1.4)⁷⁰. A concatenated multi-gene matrix was built for the CDS alignment of these 2,018 orthologous groups, which was further partitioned based on codon positions (for example, 1st, 2nd and 3rd codon positions). We used RAXML (v.8.2.6) to build the maximum likelihood tree based on the GTRGAMMA model with 100 rapid bootstraps. As an alternative, we also performed phylogenetic analysis using the consensus tree approach, in which we built individual gene trees for each of the 2,018 orthologous groups using the same method described for the concatenated tree. These individual gene trees were further summarized by ASTRAL (v.4.7.12)⁷¹ to produce the 'species tree'. Both the concatenated tree and the consensus tree were visualized in FigTree (v.1.4.2) (<http://tree.bio.ed.ac.uk/software/figtree/>).

Phenotyping. Quantitative high-throughput phenotyping was performed using end-point colony growth on solid medium. Strains were pregrown in flat-bottom 96-well microplates containing liquid YPD medium. The replicating ROTOR HDA benchtop robot (Singer Instruments) was used to mix and pin strains onto a solid YPD matrix plate at a density of 384 wells. The matrix plates were incubated

overnight at 30 °C to allow sufficient growth and replicated on 36 medium conditions, including YPD 30 °C as the pinning and growth control condition (Supplementary Table 2). Each isolate was present in quadruplicates on the corresponding matrix (interplate replicates) and at two different positions (intraplate replicates). The plates were incubated at 30 °C for 40 h and were scanned at a resolution of 600 dpi and 16-bit grayscale. Quantification of the colony size from plate images was performed using the software package Gitter⁷². Each value was normalized using the growth ratio between stress media and standard YPD medium at 30 °C (see 'Data availability'). Pairwise Pearson's correlations of fitness trait values between replicates were calculated for each condition.

Genome-wide association studies. Mixed-model association analysis was performed using FaST-LMM v.2.07¹⁹. We used the normalized phenotypes by replacing the observed value with the corresponding quantile from a standard normal distribution, as FaST-LMM expects normally distributed phenotypes. In this step, we used the markers showing a MAF > 5%. We also filtered missing genotypes as 'fs': an arbitrary threshold has been set to exclude all variants present in less than 1,000 individuals for the total matrix.

The command used for association was the following: 'fastlmm -bfile \$snp -bfileSim \$snp -pheno \$pheno -out \$assoc_file -verboseOutput'.

The mixed model adds a polygenic term to the standard linear regression designed to circumvent the effects of relatedness and population stratification. To quantify the extent of the bulk inflation and the excess false positive rate, we computed the genomic inflation factor, λ , for each condition (Supplementary Fig. 34). This factor is defined as the ratio between the median of the empirically observed distribution of the test statistic on the expected median. For example, the λ for a standard allelic test for association is based on the median (0.456) of the 1-degree-of-freedom χ^2 distribution. Under a null model of no association and unlinked variants, the expectation is for the λ to be 1. A λ greater than 1 indicates inflated P values of association, possibly owing to a confounding factor that has not been accounted for.

We estimated a trait-specific P value threshold for each condition by permuting phenotypic values between individuals 100 times. The significance threshold was the 5% quantile (the 5th lowest P value from the permutations). Using this method, variants passing this threshold have a 5% family-wise error rate (Supplementary Fig. 33).

The estimations of genome-wide heritabilities were completed by dividing the genetic variance of the null model by the total variance of the null model (genetic variance and residual variance), computed using FaST-LMM (Supplementary Fig. 6). The values reported here are based on the quantile-normalized phenotypes. To compute the variance explained by our significantly associated markers, we included them in the covariance matrix with the '-bfileSim' option and performed the same calculation again (Fig. 6).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The 1002 Yeast Genome website (<http://1002genomes.u-strasbg.fr/files/>) provides access to '-scripts.tar.gz', which contains the Perl and R scripts used to (1) extract all the non-reference material from a large set of assemblies; (2) collapse ORFs that are more similar than a specific threshold; and (3) detect regions of LOH.

Data availability. All the strains are available on request except for 11 isolates, which cannot be distributed (see Supplementary Table 1).

The Illumina reads are available in the Sequence Read Archive under accession number ERP014555.

The 1002 Yeast Genome website (<http://1002genomes.u-strasbg.fr/files/>) provides access to:

-1011Matrix.gvcf.gz: all SNPs and indels called at the population level (.gvcf format).

-1011GWASMatrix.tar.gz: the matrix used for GWAS, which contains all biallelic positions known for 1,000 isolates or more with MAF > 5% as well as CNVs (encoded 0/1/2 for absence/0.5–1 copy/multiple copies) (.bed, .bim and .fam formats).

-1011DistanceMatrixBasedOnSNPs.tab.gz: for each pair of strains, the value is the percentage, based on SNPs, of non-identical bases. Heterozygous differences were half-weighted compared to the homozygous differences.

-1011DistanceMatrixBasedOnORFs.tab.gz: for each couple of strains, the value is the number of ORFs that are present in only one out of the two isolates.

-1011Assemblies.tar.gz: de novo assemblies of the 1,011 isolates (.fasta format).
-allReferenceGenesWithSNPsAndIndelsInferred.tar.gz: sequences of the genes found in the reference genome in which SNPs and indels have been automatically inferred for each isolate.

-allORFs_pangenome.fasta.gz: sequences of the 7,796 pangenomic ORFs (.fasta format).

-genesMatrix_PresenceAbsence.tab.gz: pattern of presence and/or absence of pangenomic ORFs for each isolate, in which the presence of an ORF is marked as 1 and its absence is marked as 0.

-genesMatrix_CopyNumber.tab.gz: estimated copy number for each pangenomic ORF, per isolate. Values are given for the haploid genome, so that non-integer values can be found (different copy number on homologous chromosomes).

-genesMatrix_Frameshift.tab.gz: for each isolate, the presence or absence (indicated by 1 or 0, respectively) of homozygous frameshift is reported in each gene, based on the number of bases affected by indels.

-phenoMatrix_35ConditionsNormalizedByYPD.tab.gz: growth ratio between 35 stress conditions and standard YPD medium at 30 °C, for 971 isolates.

All other data are available from the corresponding authors upon reasonable request.

51. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* **1**, 18 (2012).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
54. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
55. Simpson, J. T. et al. ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
56. Otto, T. D., Dillon, G. P., Degraeve, W. S. & Berriman, M. RATT: rapid annotation transfer tool. *Nucleic Acids Res* **39**, e57 (2011).
57. Proux-Wéra, E., Armisen, D., Byrne, K. P. & Wolfe, K. H. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* **13**, 237 (2012).
58. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
59. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
60. Scannell, D. R. et al. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)* **1**, 11–25 (2011).
61. Baker, E. et al. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* **32**, 2818–2831 (2015).
62. Hittinger, C. T. et al. Genomics and the making of yeast biodiversity. *Curr. Opin. Genet. Dev.* **35**, 100–109 (2015).
63. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
64. Hutter, S., Vilella, A. J. & Rozas, J. Genome-wide DNA polymorphism analyses using Variscan. *BMC Bioinformatics* **7**, 409 (2006).
65. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
66. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
67. Istace, B. et al. *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* **6**, 1–13 (2017).
68. Lechner, M. et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
69. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
70. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–W612 (2006).
71. Mirarab, S. et al. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
72. Wagih, O. & Parts, L. Gitter: a robust and accurate method for quantification of colony sizes from plate images. *G3 (Bethesda)* **4**, 547–552 (2014).

Two separate outflows in the dual supermassive black hole system NGC 6240

F. Müller-Sánchez^{1*}, R. Nevin¹, J. M. Comerford¹, R. I. Davies², G. C. Privon^{3,4} & E. Treister³

Theoretical models and numerical simulations have established a framework of galaxy evolution in which galaxies merge and create dual supermassive black holes (with separations of one to ten kiloparsecs), which eventually sink into the centre of the merger remnant, emit gravitational waves and coalesce. The merger also triggers star formation and supermassive black hole growth, and gas outflows regulate the stellar content^{1–3}. Although this theoretical picture is supported by recent observations of starburst-driven and supermassive black hole-driven outflows^{4–6}, it remains unclear how these outflows interact with the interstellar medium. Furthermore, the relative contributions of star formation and black hole activity to galactic feedback remain unknown^{7–9}. Here we report observations of dual outflows in the central region of the prototypical merger NGC 6240. We find a black-hole-driven outflow of [O III] to the northeast and a starburst-driven outflow of H α to the northwest. The orientations and positions of the outflows allow us to isolate them spatially and study their properties independently. We estimate mass outflow rates of 10 and 75 solar masses per year for the H α bubble and the [O III] cone, respectively. Their combined mass outflow is comparable to the star formation rate¹⁰, suggesting that negative feedback on star formation is occurring.

NGC 6240 is an ideal system in which to study the effect of winds on the evolution of a galaxy. Its high star formation rate (SFR > 100 M_{\odot} yr^{−1}, as determined by ultraviolet and infrared luminosity measurements¹⁰; M_{\odot} , mass of the Sun) and the presence of a dual active galactic nucleus (AGN) with a separation of about 0.7 kpc (bolometric luminosities of 8×10^{44} erg s^{−1} and 2.6×10^{44} erg s^{−1} for the southwestern and northeastern nuclei, respectively^{11,12}) ensure ample activity for driving substantial feedback. A nebula with bright optical emission lines is known to be associated with NGC 6240. At scales smaller than 10 kpc, Chandra observations have revealed that the central H α nebula (the ‘butterfly nebula’) is spatially coincident with the soft X-ray-emitting gas surrounding the dual AGN^{11,13,14}. Such good spatial correlation indicates that the hot and warm phases of the interstellar medium in the butterfly nebula are excited via the same mechanism (probably shocks caused by outflowing gas^{14,15}). In addition, H α emission extends to about 90 kpc and shows multiple loops, bubbles and filaments^{14–16}. Previous studies have suggested that the extended H α nebula is primarily excited by shock heating induced by a starburst-driven superwind. However, the kinematics of the butterfly nebula has not been studied in detail because high-spatial-resolution observations are needed to separate the two nuclei and identify small-scale kinematic structures.

In this work we use Hubble Space Telescope (HST) imaging and spectroscopic data from the ground-based long-slit Dual Imaging Spectrograph at the Apache Point Observatory (APO/DIS) and the integral-field Spectrograph for Integral Field Observations in the Near-Infrared at the Very Large Telescope (VLT/SINFONI; see Methods) to determine the morphology and kinematics of the low-ionization (Br γ and H α) and high-ionization gas ([O III] at wavelength $\lambda = 5,007$ Å; hereafter, [O III]), in the butterfly-shaped inner region of NGC 6240. The HST images reveal the detailed structure of the narrow-line

region in NGC 6240 (as traced by the [O III]-emitting gas, which is less affected by star formation than H α or Br γ ^{17,18}) and allow us to ascertain whether the contribution of the dual AGN in the formation of outflowing winds is substantial.

In Fig. 1 we present a three-colour composite image of the central $25'' \times 25''$ of NGC 6240, obtained with Wide Field Camera 3 (WFC3) of the HST. Green corresponds to the continuum (filter F621M, with a central wavelength of 6,200 Å), red to H α (filter F673N) and blue to [O III] (FQ508N) emission. In the inset of Fig. 1, we summarize the relevant morphological structures. In all WFC3 images there are two emission peaks, which are located at the positions of the two AGNs, whereas most of the continuum emission is in the disk of the merger remnant. Therefore, the extended emission in the F621M image traces the stellar continuum. The prominent dark dust lane running along the northeast–southwest direction is also spatially coincident in the three WFC3 images.

The F673N image reveals filaments and bubbles of H α emission to both the east and west of the nuclei. The bubbles do not appear to be coincident with the stellar continuum, which suggests that they result from winds, rather than gas associated with tidal debris. Furthermore, the bubbles and filaments to the west of the nuclei are seen mostly in H α (regions 2 and 4). This emission is probably due to shock ionization from stellar winds (see Methods). On the other side of the nuclei, to the east/southeast of the southwestern nucleus (region 3), knots and filaments are seen in both H α and [O III], consistent with a scenario in which both the dual AGN and star formation contribute to the ionization of the extraplanar gas. In addition to region 3, the FQ508N image also reveals diffuse [O III] emission in some parts of the galactic disk and an extended extraplanar structure with conical geometry to the northeast of the nuclei (region 1), which is faint to non-existent in H α . The [O III] gas in region 1 is not spatially correlated with the stellar continuum (Fig. 1) or with the molecular gas (Extended Data Fig. 2), which indicates that this structure is not associated with tidal tails from the merger. These four observations, as well as the similarity between the morphology of the [O III] emission (which extends to a distance of 3.7 kpc to the northeast with an opening angle $\theta_{\text{out}} \approx 50^\circ$; see Extended Data Fig. 1) with ionization cones seen in prototypical Seyfert galaxies^{17,18}, suggest that the gas in region 1 is mostly ionized by the AGN. This interpretation is supported by optical emission-line diagnostics (see Methods). Most of the points in the [O III] cone lie in the Seyfert or LINER (low-ionization nuclear-emission-line region), rather than in the starburst (H II) region, on the Baldwin–Phillips–Terlevich (BPT) diagram¹⁹ (Extended Data Fig. 3), with 20% (3 out of 15) of the spatial elements located in the Seyfert region. In addition, the [O III]/H β ratio is considerably enhanced in region 1, reaching a maximum of about 10 (the largest in the butterfly nebula), placing the [O III] cone firmly in the Seyfert region^{19–21}. This behaviour is not observed in any other part of the galaxy and suggests a substantial contribution from AGN photoionization. Because LINERs do not usually show ionization cones of highly ionized gas^{22,23}, the [O III] cone is probably produced by the southwestern nucleus, which is also the more powerful of the

¹Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA. ²Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany. ³Instituto de Astrofísica and Centro de Astroingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, Santiago, Chile. ⁴Department of Astronomy, University of Florida, Gainesville, FL, USA.

*e-mail: francisco.mullersanchez@colorado.edu

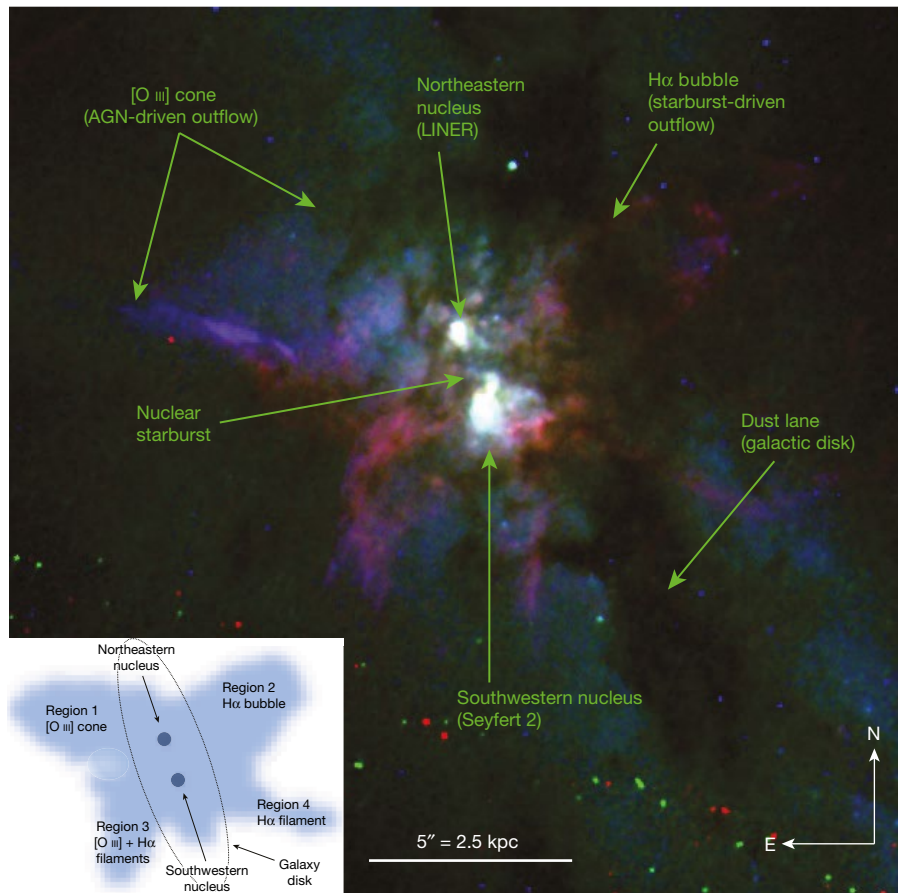


Fig. 1 | Three-colour composite image of H α (red), [O III] (blue) and V-band continuum (green) emission in NGC 6240, obtained by the HST. The major structures identified in the central $25'' \times 25''$ of NGC 6240 are labelled. The inset shows the morphological structures identified in the butterfly-shaped region of NGC 6240 using HST narrow-band filters. The northeastern nucleus shows a typical LINER spectrum and is about three times fainter than the southwestern nucleus, which exhibits the

characteristics of a heavily obscured Seyfert 2 galaxy^{11,12}. The ionized gas in regions 1–4 is extraplanar. The [O III] cone extends to about 4 kpc to the northeast (which is faint in H α), which indicates an AGN-driven outflow, whereas the H α bubble to the northwest is indicative of a starburst-driven outflow. All physical units in this paper are based on a concordance, flat Λ CDM cosmology with a Hubble constant of $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. At the redshift of NGC 6240 ($z = 0.0245$), $1''$ corresponds to about 0.5 kpc.

two AGNs in the system. However, we cannot rule out a contribution from the northeastern nucleus. The location of the [O III] cone (next to the northeastern nucleus) and the fact that the apparent base of the cone encompasses both nuclei suggest a combined narrow-line region from the two AGNs.

Figure 2 shows the SINFONI velocity map of ionized hydrogen (Br γ) overlaid on the H α HST image. The velocity maps of the stars and molecular hydrogen (H $_2$) are also shown for comparison. While two rotational components are observed in the stellar kinematics, the H $_2$ velocity map shows one large perturbed rotating disk with a rotation axis that is not aligned with those of the two nuclei. The decoupled H $_2$ disk is probably produced by the tidal forces of the interaction and its sense of rotation follows the orbital history of the merger^{24,25}. The rotational velocity of the H $_2$ disk is about 220 km s^{-1} . Interestingly, the kinematics of the low-ionization gas is considerably different from both that of stars and that of H $_2$. Redshifted velocities of 360 km s^{-1} and broad emission lines (velocity dispersion $\sigma \approx 450 \text{ km s}^{-1}$) are observed in the northeastern nucleus at the base of the H α bubble (Extended Data Fig. 4), indicating an outflow of ionized hydrogen (see Methods).

Position–velocity diagrams of the [O III] emission in region 1 reveal the typical signatures of outflows (Fig. 3): (i) high-velocity (up to 350 km s^{-1}) components that cannot be explained by the same gravitational potential that produces H $_2$ velocities of about 220 km s^{-1} in the galaxy disk (Fig. 2), (ii) broad components of [O III] ($\sigma \approx 1,070 \text{ km s}^{-1}$) and (iii) signatures of radial acceleration and deceleration^{5,18,26}. These features support the hypothesis of a non-gravitational force accelerating the gas from about 0 km s^{-1} at the centre of the galaxy (between the

two nuclei; see Extended Data Fig. 1) to about 350 km s^{-1} at a distance $r = 1.8 \text{ kpc}$ and subsequently decelerating it to about 190 km s^{-1} at $r = 3.7 \text{ kpc}$ (Fig. 3).

The morphology, kinematics, timescale and energetics of the [O III] cone are consistent with energy injection from the AGN (see Methods). In particular, the kinetic power of the outflow is about 2.9 times larger than the estimated injection of energy from the nuclear starburst (assuming $\text{SFR} = 100 M_{\odot} \text{ yr}^{-1}$), requiring a substantial contribution from the AGN. On the other hand, the timescale of the H α bubble (7.4 Myr) and its energetics are consistent with energy injection from a recent episode of star formation that started less than 9 Myr ago. This timescale is inconsistent with the typical AGN flickering cycles^{27–29} but agrees with the age of the nuclear starburst in NGC 6240^{14,25,30}.

The AGN-driven outflow carries about 7.5 times more mass ($\dot{M}_{\text{AGN}} \approx 75 M_{\odot} \text{ yr}^{-1}$) and is about 15 times more powerful (higher kinetic luminosity, $\dot{E}_{\text{AGN}} \approx 2 \times 10^{44} \text{ ergs s}^{-1}$) than the outflow in the H α bubble ($\dot{M}_{\text{bubble}} \approx 10 M_{\odot} \text{ yr}^{-1}$, $\dot{E}_{\text{bubble}} \approx 1.3 \times 10^{43} \text{ ergs s}^{-1}$). We note that \dot{M}_{bubble} does not correspond to the total outflow rate due to star formation (\dot{M}_{sf}) in the nuclear region; regions 3 and 4 also need to be included. Assuming the same properties of the H α bubble (geometry, kinematics and mass) for regions 3 and 4 (Fig. 1; see also refs¹⁴ and ¹⁵), \dot{M}_{sf} would be about $30 M_{\odot} \text{ yr}^{-1}$. We use the ratio of the mass outflow rate to the SFR to evaluate the influence of negative feedback on the newly formed galaxy disk⁶. A value smaller than 1 indicates that the outflow does not carry enough mass to affect the stellar production considerably. If this ratio is equal to or greater than 1, negative feedback on star formation is occurring. In NGC 6240 the combined effect of

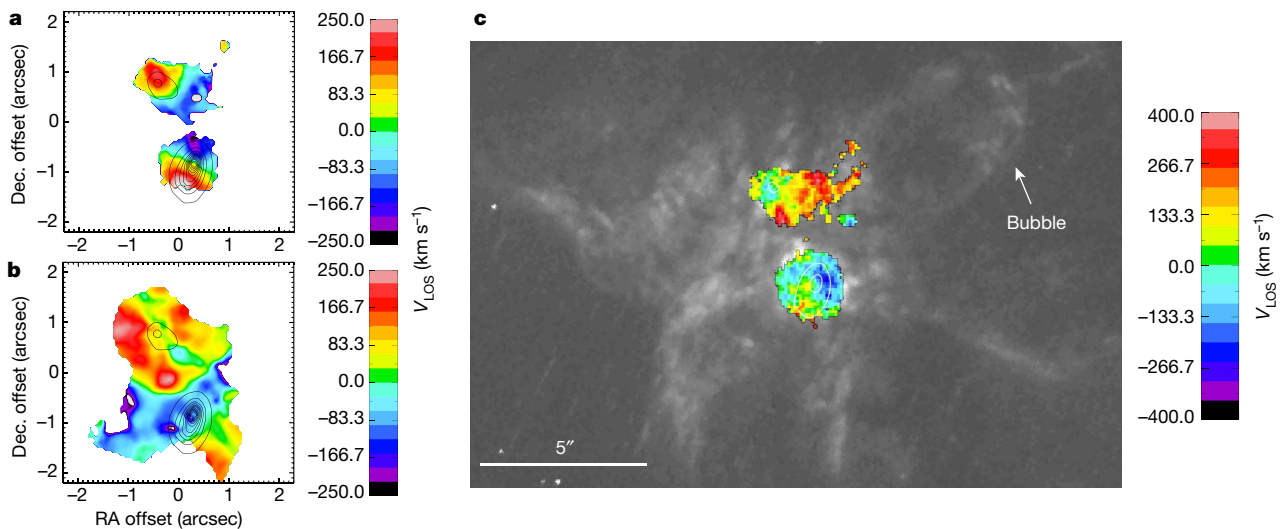


Fig. 2 | VLT/SINFONI maps of NGC 6240. **a**, Stellar velocity. **b**, H_2 velocity map. **c**, Velocity map of $\text{Br}\gamma$ overlaid on the HST image of $\text{H}\alpha$. In all panels, the contours delineate the K-band continuum emission and are spaced at 10% of the peak flux. Although two rotational components are observed in the stellar kinematics, the H_2 kinematics shows one perturbed rotational component with a kinematic major axis oriented at a position

angle of 22° . The $\text{Br}\gamma$ kinematics of the northeastern nucleus is dominated by non-circular motions. Redshifted velocities of 360 km s^{-1} are observed at the base of the $\text{H}\alpha$ bubble. In all maps, north is up and east is to the left. RA, right ascension; Dec., declination. The colour bar indicates line-of-sight velocity (V_{LOS}).

\dot{M}_{AGN} and \dot{M}_{sf} is comparable to the SFR¹⁰. Therefore, we are witnessing the crucial phase in the evolution of mergers of gas-rich galaxies in which suppression of star formation is starting to occur. It is important

to note that the starburst-driven outflow alone underestimates the effect of feedback in the galaxy (a similar conclusion is reached for the AGN-driven outflow). Only the combined mass outflow rate can limit

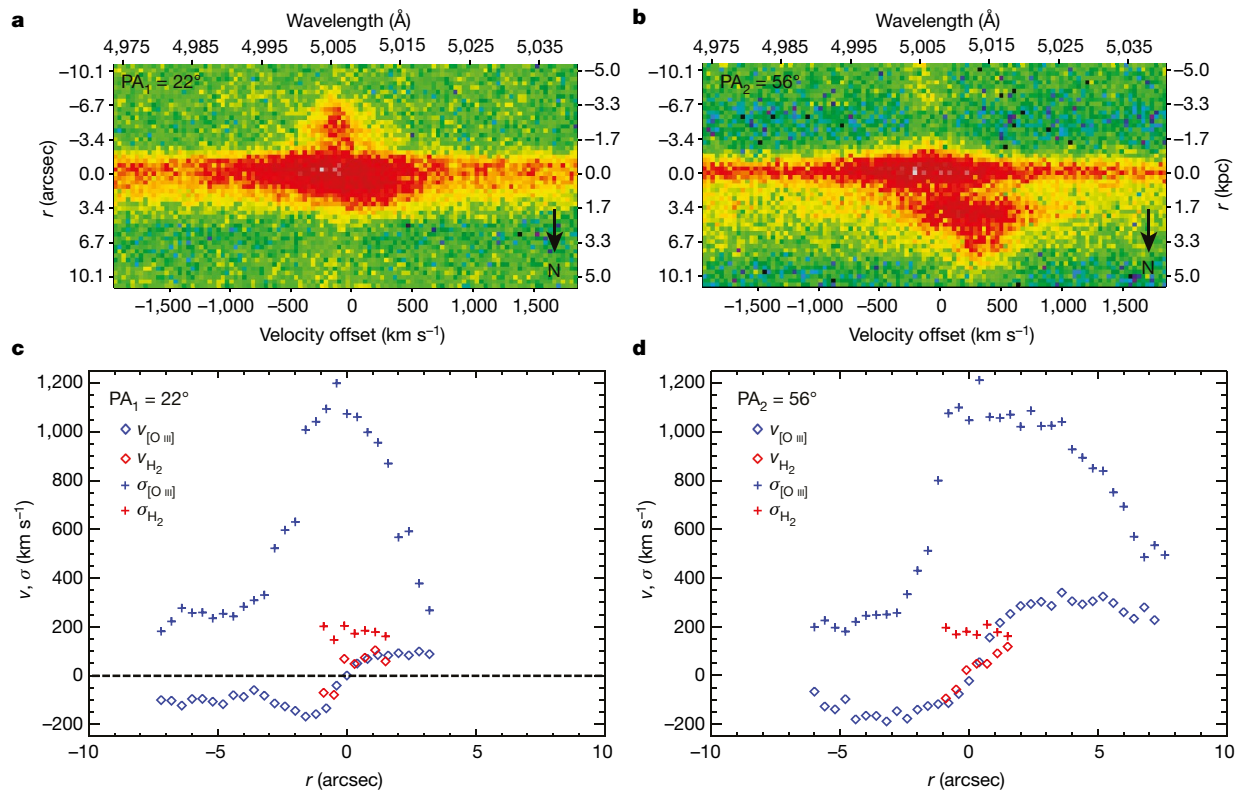


Fig. 3 | Kinematics of [O III]. **a**, **b**, Segments of the two-dimensional long-slit spectra of NGC 6240, centred at the rest wavelength of [O III]. The colour scale represents flux density normalized to the peak. Cool colours (green and blue) correspond to background emission (<10% of the peak emission). Warm colours (yellow to red) correspond to sizeable flux density values (>10% of the peak of emission). **c**, **d**, Position-velocity diagrams of [O III] and H_2 emission, where v is the velocity and σ is the velocity dispersion. The galaxy was observed at two position angles, 22° and 56° (see also Extended Data Fig. 1). Positive values of angular distance

(vertical axis in **a** and **b** and horizontal axis in **c** and **d**) correspond to the direction north from the centre of the galaxy, at the marked position angle. The number of spatial elements extracted from our long-slit observations (**a**, **b**) is 27 at $\text{PA}_1 = 22^\circ$ (**c**) and 34 at $\text{PA}_2 = 56^\circ$ (**d**). There are 15 spatial elements inside the [O III] cone (between $0'' < r < 7''$ in **d**). We extracted the velocity and dispersion values for H_2 at 7 different spatial positions along imaginary APO/DIS long slits oriented at 22° and 56° in the SINFONI data (Fig. 2).

the star formation activity and the growth of the newly formed galaxy after the merger event.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0033-2>.

Received: 5 July 2017; Accepted: 4 January 2018;

Published online 18 April 2018.

- Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433**, 604–607 (2005).
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E. & Hernquist, L. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **391**, 481–506 (2008).
- Hopkins, P. F., Quataert, E. & Murray, N. Stellar feedback in galaxies and the origin of galaxy-scale winds. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
- Veilleux, S., Cecil, G. & Bland-Hawthorn, J. Galactic winds. *Annu. Rev. Astron. Astrophys.* **43**, 769–826 (2005).
- Müller-Sánchez, F. et al. Outflows from active galactic nuclei: kinematics of the narrow-line and coronal-line regions in Seyfert galaxies. *Astrophys. J.* **739**, 69 (2011).
- Bolatto, A. D. et al. Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind. *Nature* **499**, 450–453 (2013).
- Heckman, T. M. & Best, P. N. The coevolution of galaxies and supermassive black holes: insights from surveys of the contemporary Universe. *Annu. Rev. Astron. Astrophys.* **52**, 589–660 (2014).
- Karouzos, M. et al. A tale of two feedbacks: star formation in the host galaxies of radio AGNs. *Astrophys. J.* **784**, 137 (2014).
- Roos, O., Juneau, S., Bournaud, F. & Gabor, J. M. Thermal and radiative active galactic nucleus feedback have a limited impact on star formation in high-redshift galaxies. *Astrophys. J.* **800**, 19 (2015).
- Howell, J. H. et al. The great observatories all-sky LIRG survey: comparison of ultraviolet and far-infrared properties. *Astrophys. J.* **715**, 572–588 (2010).
- Komossa, S. et al. Discovery of a binary active galactic nucleus in the ultraluminous infrared galaxy NGC 6240 using Chandra. *Astrophys. J.* **582**, L15–L19 (2003).
- Puccetti, S. et al. Hard X-ray emission of the luminous infrared galaxy NGC 6240 as observed by NuSTAR. *Astron. Astrophys.* **585**, A157 (2016).
- Lira, P., Ward, M. J., Zezas, A. & Murray, S. S. Chandra HRC and HST observations of NGC 6240: resolving the active galactic nucleus and starburst. *Mon. Not. R. Astron. Soc.* **333**, 709–714 (2002).
- Yoshida, M. et al. Giant H α nebula surrounding the starburst merger NGC 6240. *Astrophys. J.* **820**, 48 (2016).
- Heckman, T. M., Armus, L. & Miley, G. K. On the nature and implications of starburst-driven galactic superwinds. *Astrophys. J.* **74**, 833–868 (1990).
- Veilleux, S., Shopbell, P. L., Rupke, D. S., Bland-Hawthorn, J. & Cecil, G. A search for very extended ionized gas in nearby starburst and active galaxies. *Astrophys. J.* **126**, 2185–2208 (2003).
- Schmitt, H., Donley, J. L., Antonucci, R. R. J., Hutchings, J. B. & Kinney, A. L. A Hubble space telescope survey of extended [O III] 5007 emission in a far-infrared selected sample of Seyfert galaxies: observations. *Astrophys. J.* **148**, 327–352 (2003).
- Fischer, T. C., Crenshaw, D. M., Kraemer, S. B. & Schmitt, H. R. Determining inclinations of active galactic nuclei via their narrow-line region kinematics. I. Observational results. *Astrophys. J.* **209**, 1 (2013).
- Baldwin, J. A., Phillips, M. M. & Terlevich, R. Classification parameters for the emission-line spectra of extragalactic objects. *Publ. Astron. Soc. Pacif.* **93**, 5–19 (1981).
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A. & Trevena, J. Theoretical modeling of starburst galaxies. *Astrophys. J.* **556**, 121–140 (2001).
- Kewley, L. J., Groves, B., Kauffmann, G. & Heckman, T. The host galaxies and classification of active galactic nuclei. *Mon. Not. R. Astron. Soc.* **372**, 961–976 (2006).
- Pogge, R. W., Maoz, D., Ho, L. C. & Eracleous, M. The narrow-line regions of LINERS as resolved with the Hubble space telescope. *Astrophys. J.* **532**, 323–339 (2000).
- Müller-Sánchez, F. et al. The central molecular gas structure in LINERs with low-luminosity active galactic nuclei: evidence for gradual disappearance of the torus. *Astrophys. J.* **763**, L1 (2013).
- Tacconi, L. J. et al. Gas dynamics in the luminous merger NGC 6240. *Astrophys. J.* **524**, 732–745 (1999).
- Tecza, M. et al. Stellar dynamics and the implications on the merger evolution in NGC 6240. *Astrophys. J.* **537**, 178–190 (2000).
- Crenshaw, M. & Kraemer, S. B. resolved spectroscopy of the narrow-line region in NGC 1068: kinematics of the ionized gas. *Astrophys. J.* **532**, L101–L104 (2000).
- Davies, R. I. et al. Fueling active galactic nuclei. II. Spatially resolved molecular inflows and outflows. *Astrophys. J.* **792**, 101 (2014).
- Hickox, R. C. et al. Black hole variability and the star formation–active galactic nucleus connection: do all star-forming galaxies host an active galactic nucleus? *Astrophys. J.* **782**, 9 (2014).
- Schawinski, K., Koss, M., Berner, S. & Sartori, L. F. Active galactic nuclei flicker: an observational estimate of the duration of black hole growth phases of $\sim 10^5$ yr. *Mon. Not. R. Astron. Soc.* **451**, 2517–2523 (2015).
- Engel, H. et al. NGC 6240: merger-induced star formation and gas dynamics. *Astron. Astrophys.* **524**, A56 (2010).

Acknowledgements Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). The Space Telescope Science Institute is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. The optical spectroscopic data reported here were obtained at the Apache Point Observatory 3.5-m telescope, which is owned and operated by the Astrophysical Research Consortium. F.M.-S. acknowledges financial support from NASA HST Grant HST-AR-13260.001. G.C.P. acknowledges support from a FONDECYT Postdoctoral Fellowship (number 3150361) and the University of Florida. E.T. acknowledges support from CONICYT Anillo ACT1101, FONDECYT regular grants 1120061 and 1160999, and Basal-CATA PFB-06/2007.

Author contributions F.M.-S. conceived the project, analysed the data, coordinated the activities and prepared the manuscript. R.N. prepared and reduced the APO/DIS observations and created the BPT diagrams. F.M.-S. and J.C. analysed the HST images. R.D. reduced the VLT/SINFONI data. E.T. and G.C.P. contributed to the analyses and discussion. All authors discussed the results and implications and commented on the manuscript at all stages.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0033-2>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to F.M.-S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

HST imaging. NGC 6240 was observed with the WFC3 on 2 August 2012 (programme GO-12552; principal investigator, L. Kewley). The HST/WFC3 data provide high-resolution ($0.0396''$ per pixel) optical imaging in five narrow-band and two medium-band filters (WFC3-UVIS channel). In this work, we selected the filters FQ508N and F673N as tracers of the high-ionization ([O III]) and low-ionization (H α) gas, respectively. At the redshift of NGC 6240 ($z = 0.0245$), these two filters entirely cover the [O III] and H α emission. Blue (F467M) and red (F621M) continuum images were also used to trace the emission of the stars in the system. We used continuum-subtracted [O III] and H α images for the analysis. To obtain an image of the ionized gas emission ([O III] and H α), we aligned the off-band image with the on-band one, scaled it to fit the on-band image at radii where the ionized gas flux is negligible, and subtracted it from the on-band image.

Near-infrared adaptive-optics-assisted integral-field spectroscopy. The integral-field data of NGC 6240 used in this work are VLT/SINFONI observations^{31,32} obtained on the night of 20 August 2007 (programme 079.B-0576). Details of the observations and data reduction have been described in a previous publication³⁰. The final datacube has a spatial resolution of $0.097'' \times 0.162''$ full-width at half-maximum (using a pixel scale of $0.05''$ per pixel and a field of view of $3.6'' \times 4.0''$) and a spectral resolution of about 90 km s^{-1} full-width at half-maximum in the K-band. We derived the two-dimensional properties (flux and velocity maps) of the gas and the stars using the IDL code LINEFIT³³, which estimates the uncertainties in the Gaussian fits using Monte Carlo techniques. This method fits the emission lines (absorption features) by convolving a Gaussian with a spectrally unresolved template profile (a sky line for emission lines and a template stellar spectrum for stellar absorption features) to the continuum-subtracted spectral profile at each spatial pixel in the datacube. The velocities are measured relative to the systemic redshift of the galaxy, taking into account in the stellar kinematics the 50 km s^{-1} redshift of the northeastern nucleus with respect to the southwestern nucleus^{25,30}. Several realizations (usually 100) of the data are generated by adding random noise to the flux at each pixel; these are fitted using the same procedure as above. This method allowed us to obtain uncertainties for the kinematic maps in the range $30\text{--}40 \text{ km s}^{-1}$. We note that the near-infrared spectrum of NGC 6240 does not show any strong high-ionization lines from the narrow-line region (the coronal lines [Si VI] and [Ca VIII] are not present in NGC 6240³⁰), and therefore a new optical spectrum was acquired.

Optical long-slit spectroscopy. The long-slit spectroscopic observations of NGC 6240 were obtained on 30 June 2016 using the DIS with a $1.5'' \times 6'$ slit at the APO. We used a grating with 1,200 lines per millimetre to obtain a spatial resolution of 0.62 \AA per pixel. The spatial scales were $0.4''$ per pixel and $0.42''$ per pixel in the red and blue channels, respectively. We observed with two slit positions, for twenty minutes at each position. Slit position 1 (PA₁ hereafter) was oriented at 22° (the position angle is measured counter-clockwise, from north, 0° , to east, 90°), whereas slit position 2 (PA₂ hereafter) was oriented at 56° (see Extended Data Fig. 1). The emission lines were modelled as Gaussians. For each row in the spatial direction, we fitted a single Gaussian combined with a straight line with a given slope (representing the continuum emission) to each emission line. The uncertainties were estimated using Monte Carlo techniques. The method involves adding noise to the galaxy spectra and refitting the result (about 500 times) to empirically determine the standard deviation of each fitted parameter. Using this method we obtained uncertainties of $70\text{--}160 \text{ km s}^{-1}$ for the velocity and dispersion measurements and flux errors of up to 25%.

Emission-line diagnostics. In Extended Data Fig. 3 we show the BPT¹⁹ emission-line diagnostic diagrams for all the spatial positions along the two slits where the relevant emission lines—H β , [O III], H α and [N II] at $\lambda = 6,584 \text{ \AA}$ (hereafter, [N II])—have a signal-to-noise ratio greater than 3. To construct the BPT diagrams for each position angle, we measure emission-line fluxes at all spatial positions of the galaxy (for H β , [O III], H α and [N II]). We first determine the spatial centre of the galaxy from the galaxy continuum (Fig. 3) and then fit a single Gaussian for H β and [O III]. We fit three Gaussians to the H α –[N II] complex, where we require the velocity dispersions of the [N II] lines to be identical and their flux ratio to be 1:3.

Most of the emission in the galaxy disk (PA₁ = 22°) is consistent with H II/LINER excitation. Two data points are located on the border between the star-forming and the Seyfert regions of the diagram, but these could be artefacts caused by a low signal-to-noise ratio (weak [O III] and H β emission is observed in the north part of the galaxy disk; Figs. 1 and 3). Along the direction of PA₂ = 56° , the situation is different. At distances between $-6''$ and $-1''$ from the nucleus (distances are negative to the south), the emission is located in the star-forming part of the BPT diagram. Between $-1''$ and $1''$ (the centre of the galaxy), the [N II]/H α ratio increases from about 0.3 to about 5, shifting data points towards the LINER region of the diagram. Finally, inside the [O III] cone, at distances between $1''$ and $6''$, [O III]/H β increases from about 2 to about 10, shifting data points towards the Seyfert region of the diagram. However, because the [N II]/H α ratios are also high (> 2), some points remain in the LINER region. High [N II]/H α values

(> 1.5) usually correspond to LINER-like excitation from shocks^{34–36}. Therefore, the region occupied by the [O III] cone in the BPT diagram suggests the presence of strong shocks and AGN photoionization.

From our SINFONI data, we also obtained a map of the H $_2$ /Br γ ratio (Extended Data Fig. 5). This map is usually used as a tracer of shocks in the near-infrared. If H $_2$ is excited by ultraviolet photons from stars, this ratio should be small because the photons from those stars should also produce substantial Br γ emission (star-forming regions have H $_2$ /Br γ ratios^{37–39} < 0.6). A high H $_2$ /Br γ value (> 1) would indicate shocks rather than star formation^{36,40}. As can be seen in Extended Data Fig. 5, our data show very high H $_2$ /Br γ ratios (3–48), with the maximum of 48 located between the two nuclei. Therefore, the bright H $_2$ emission in the central region of NGC 6240 is produced by shocks. The large H $_2$ /Br γ ratio of 48 (the largest value in the local Universe) is probably produced⁴¹ by the collision of the interstellar media associated with the two progenitor galaxies. In addition, the combination of moderately high H $_2$ /Br γ ratios (1–15) with high velocity dispersion values ($\sigma > 250 \text{ km s}^{-1}$) is usually associated with outflows^{35,36,41}. These two characteristics are observed at the inner edge (or base) of the H α bubble (Extended Data Figs 4 and 5), suggesting the presence of an outflowing wind in this region.

Our results confirm previous studies of the optical and near-infrared emission-line ratios in NGC 6240, which suggest that the emission in the central $3.0'' \times 3.0''$ region of the galaxy is dominated by shocks, showing a typical shock-excited LINER spectrum. The near-infrared data imply the presence of two types of shocks: (i) strong shocks, with high H $_2$ /Br γ ratios (higher than 15), between the nuclei, caused by the collision of the interstellar media of the merging galaxies and (ii) shocks with slightly lower H $_2$ /Br γ ratios (3–15) and higher velocity dispersions (up to about 450 km s^{-1}), produced by outflowing winds. The inner edge of the H α bubble is consistent with the latter type of shocks. Our APO/DIS observations provide information on the excitation mechanisms of the ionized gas in the [O III] cone. Our BPT diagrams suggest the presence of shocks and AGN photoionization in this region. Finally, our long-slit data indicate that the nebular emission in region 4 is dominated by star formation.

Evidence of outflows in the H α bubble. Several pieces of evidence suggest that the kinematics of the H α bubble is dominated by outflows. First, the ionized gas (Br γ) in the northeastern nucleus exhibits line-of-sight velocities of 360 km s^{-1} , which are too high to be explained by the same gravitational potential that is producing rotational star velocities of 200 km s^{-1} in this nucleus (Fig. 2). Furthermore, the velocity map of Br γ is very different to that of the stars. The kinematic major axis of the stars has a position angle of about 37° , which is consistent with the photometric major axis of this nucleus (obtained from adaptive-optics images^{30,42}). The kinematic major axis of Br γ extends in the east–west direction with a particularly fast component to the northwest (360 km s^{-1}), which is spatially coincident with the H α bubble seen in the HST images. A similar result is obtained when comparing the Br γ velocity map with that of H $_2$. The molecular gas disk has a rotation axis that is not aligned with that of either nucleus^{24,25}. The H $_2$ disk has a position angle of 22° and a redshifted velocity of $220 \pm 33 \text{ km s}^{-1}$ in the northeastern nucleus. The velocity map of Br γ shows both redshifted and blueshifted velocities in this nucleus. In addition, the maximum Br γ velocity of 360 km s^{-1} is inconsistent with the maximum rotational velocity of the molecular gas in the disk of the advanced merger (about 220 km s^{-1}). Finally, the gas at the base of the H α bubble has a very high velocity dispersion ($450 \pm 67 \text{ km s}^{-1}$), which can be explained only by outflows at these scales^{4,5,35,43}. All these results support strongly the premise of a non-gravitational force that is accelerating the gas in the northeastern nucleus to a maximum line-of-sight velocity of about 360 km s^{-1} at $r = 3.3''$ (1.6 kpc) with a maximum velocity dispersion of about 450 km s^{-1} . These results are broadly consistent with those of a previous study¹⁵, which found a maximum line-of-sight velocity of about 400 km s^{-1} for the region covered by the H α bubble in seeing-limited long-slit observations of NGC 6240.

Evidence for outflows in the [O III] cone. Two structural features suggest strongly the presence of outflows in the [O III] cone: (i) the location of the [O III] cone is outside the plane of rotation of the disk of the merger, in a region that is not associated with tidal structures from the galaxy merger (see Extended Data Fig. 2), and (ii) its morphology is typical of outflows seen in prototypical starburst and Seyfert galaxies^{5,6,8,44}. Conical morphologies are not expected for inflows, which are usually radial streamers of gas^{45,46}. These two characteristics rule out rotation or inflows as possible kinematic components of the [O III] emission in region 1.

The kinematics of the [O III] cone indicates the presence of outflows. Figure 3 shows the two-dimensional APO/DIS spectra of NGC 6240 and the kinematics (position–velocity diagrams) of [O III], extracted at the two position angles (PA₁ and PA₂) indicated in Extended Data Fig. 1. For comparison, we also extracted the kinematics of H $_2$ along the directions of PA₁ and PA₂, matching the width of the optical slits. At both position angles, the spectra exhibit a broad [O III] component ($\sigma = 1,220 \pm 140 \text{ km s}^{-1}$) in the central $1.5''$, with a line-of-sight velocity consistent with the systemic velocity of the galaxy. These extremely broad lines suggest the existence of an outflow that originates from the central $1.5''$ of the galaxy

(the width of the long slit of the DIS). In addition, one clear trend emerges: the gas is more kinematically disturbed along the direction of the morphologically inferred outflow ($PA_2 = 56^\circ$) than along the disk of the merger ($PA_1 = 22^\circ$). At $PA_1 = 22^\circ$, the [O III] emission is blueshifted by about 180 km s^{-1} in the south, redshifted by about 100 km s^{-1} in the north, and the emission lines are narrow ($\sigma < 250 \text{ km s}^{-1}$), consistent with the curves of H_2 .

By contrast, at $PA_2 = 56^\circ$, the redshifted broad emission lines to the north provide support for an outflow-dominated kinematic component. As can be seen in Fig. 3, the [O III] velocity curve deviates considerably from that of H_2 at distances $r > 1''$. The H_2 velocity reaches a maximum of 140 km s^{-1} at $r = 2''$. At this distance, the [O III] velocity is about 250 km s^{-1} , and it continues to increase with distance, reaching a maximum velocity of $v_{\text{max}} = 350 \pm 30 \text{ km s}^{-1}$ at $r = 3.6''$. This velocity is too high to be explained by the gravitational potential of the H_2 disk. The maximum dispersion of [O III] is $1,070 \pm 110 \text{ km s}^{-1}$ at $r = 2.1''$. We adopted this value as the representative velocity dispersion of the nebula because the velocity $1,220 \text{ km s}^{-1}$ at $r \approx 0''$ might be affected by other random motions caused by the merger process (see Extended Data Fig. 5). The outflow component then begins to decelerate outside this maximum-velocity region, reaching about 190 km s^{-1} at $r = 7.4''$, which is an observational signature of an outflow encountering drag forces from the interstellar medium^{5,26}. In the south region, along the direction of PA_2 , the [O III] emission is faint and narrow, with line-of-sight velocities consistent with the rotation of the disk of the advanced merger (Fig. 3). Finally, the broad kinematic components in combination with the high [N II]/H α ratios (characteristic of shock ionization; see Extended Data Fig. 3), provide strong evidence for the existence of outflowing gas in the [O III] cone^{35,36}.

Estimation of mass outflow rates. The amount of feedback, in terms of outflowing mass entrained in the [O III] cone and H α bubble, can be estimated using the morphological parameters derived from the HST images and the velocities measured in our APO/DIS and VLT/SINFONI data. For the [O III] cone, we estimate the mass outflow rate (\dot{M}_{cone}) using a method described in an earlier publication⁵ and by assuming a gas density of $n_e = 50 \text{ cm}^{-3}$ and a filling factor of $f = 0.01$, which are typical of the narrow-line region at $r \approx 1.5 \text{ kpc}^{14,15,47}$. The [O III] cone has an opening angle of $50 \pm 3^\circ$ (Extended Data Fig. 1), and the projected distance at which the outflow reaches $v_{\text{max}} = 350 \text{ km s}^{-1}$ is $r_t = 1.8 \text{ kpc}$. At $r > r_t$ the deceleration phase starts (see Fig. 3d). We have derived the characteristic outflow speed as $v_{\text{out}} = \sqrt{v_{\text{max}}^2 + \sigma_{\text{mean}}^2} / \sin i$, where i is the inclination and $\sigma_{\text{mean}} = 505 \pm 120 \text{ km s}^{-1}$ is the average velocity dispersion of the gas inside the [O III] cone (Fig. 3). This term takes into account the spread of velocities along the line of sight and additional turbulence that may be substantial. Although our geometric model of the [O III] cone does not constrain i , it excludes values smaller than 25° (which would imply that the face of the cone that is closer to us is blueshifted) and greater than 65° (a nearly face-on view to the cone). We assume that $\sin i \approx \cos i \approx 0.71$ to account for the unknown inclination of the outflowing gas (this correction is also applied to the distance at which the outflow reaches its maximum velocity r_t). The resulting mass outflow rate is $\dot{M}_{\text{cone}} = 75 M_\odot \text{ yr}^{-1}$, with an uncertainty of a factor of three. We were able to mitigate the uncertainties in the geometrical parameters and kinematics of the [O III] cone thanks to the high-resolution HST images and our detailed curves of line-of-sight velocity and velocity dispersion as a function of position. The uncertainty of a factor of three in \dot{M}_{cone} comes from the assumed value of density, which for NGC 6240¹⁵ and other galaxies with high infrared luminosities is estimated to be in the range $20\text{--}150 \text{ cm}^{-3}$, with a typical value of about 50 cm^{-3} at $r \approx 2 \text{ kpc}$ (in other words, $\log n_e = 1.7 \pm 0.47$). We point out that the obtained value of \dot{M}_{cone} is probably a conservative estimate for the mass outflow rate because we may have underestimated the outflow covering factor (although we considered only the region inside the conical structure of [O III] emission, the outflow might be covering a larger volume) and the maximum outflow velocity (up to about $1,070 \text{ km s}^{-1}$ for the high-dispersion [O III] clouds).

For the H α bubble, we calculated the outflow rate²⁷ as $\dot{M}_{\text{bubble}} = M_{\text{H}\alpha} v_{\text{out}} / r$ because the mass of ionized hydrogen is known. A mass of $1.4 \times 10^6 M_\odot$ has been estimated¹⁴ for the butterfly-shaped nebula, assuming a spherically symmetric structure. The H α bubble covers approximately one quadrant of the butterfly nebula (Fig. 1). Therefore, the mass of ionized hydrogen in this region is about $3.5 \times 10^7 M_\odot$. The maximum velocity of the gas is 360 km s^{-1} at the inner edge of the bubble (Fig. 2) and $\sigma_{\text{mean}} = 260 \pm 40 \text{ km s}^{-1}$ (Extended Data Fig. 4). The H α bubble appears to be perpendicular to the galaxy disk and the northeastern nucleus, which has an inclination^{25,30} of about 45° . Therefore, we again adopt $\sin i \approx \cos i \approx 0.71$ as a de-projection factor. Assuming the bubble is expanding at a constant velocity up to a radius of 1.6 kpc , then $\dot{M}_{\text{bubble}} = 10 M_\odot \text{ yr}^{-1}$. The largest uncertainty in the calculation of \dot{M}_{bubble} comes from the mass of ionized gas, which has been estimated¹⁴ as 20%. However, we are probably underestimating the mass of H α in the H α bubble, which can be up to three times larger (taking into account all the hydrogen gas that is being ionized by star formation in regions 3 and 4).

Driving mechanism of the outflows. Here we compare the morphologies, velocities, timescales and energetics of the [O III] cone and the H α bubble to identify the

primary driver of the outflow in each region. We calculate the dynamical time of the outflows as $t_{\text{dyn}} = D/v_{\text{out}}$, where D is the size of the [O III] cone or H α bubble, and obtain $7.4 \pm 1.4 \text{ Myr}$ for the H α bubble and $3.9 \pm 1.2 \text{ Myr}$ for the [O III] cone. The timescale of the H α bubble is inconsistent with the typical timescale of the active phase of an AGN (the AGN can flicker on and off, showing a variability of 1–3 orders of magnitude, with a timescale of 0.1 Myr to a few million years owing to stochastic accretion at small scales^{27–29}), but agrees with the age of the most recent starburst in NGC 6240 ($6\text{--}9 \text{ Myr}$)^{14,25,30}. By contrast, the timescale of the [O III] cone is similar to those of typical AGN flickering cycles²⁹, suggesting the presence of an AGN-driven outflow in region 1 of the butterfly nebula.

Next, we estimate the amount of energy injection required to power the outflows (kinetic power and luminosity). We use two methods⁴³: the first method⁵ estimates the kinetic power of the outflow as $\dot{E} = \dot{M} v_{\text{out}}^2 / 2$. The second method treats the outflows as analogous to supernova remnants, but with continuous energy injection, and estimates the energy injection rate required to expand the outflows into a low-density medium. Using the first method we obtain kinetic powers of $1.2 \times 10^{42} \text{ erg s}^{-1}$ for the H α bubble and $1.9 \times 10^{43} \text{ erg s}^{-1}$ for the [O III] cone. As mentioned earlier, these are conservative values (probably lower limits) for the kinetic power of the outflows because we are probably underestimating the mass outflow rates. For the second method, we used equation (2) in ref. ⁴³ with an ambient density of a uniform low-density medium $n_0 = 0.5 \text{ cm}^{-3}$ and a covering factor of 1 (these values probably represent upper limits for these parameters). We found an energy injection rate of $1.5 \times 10^{44} \text{ erg s}^{-1}$ and $2 \times 10^{45} \text{ erg s}^{-1}$ for the H α bubble and the [O III] cone, respectively. We adopted a fiducial range between $1.2 \times 10^{42} \text{ erg s}^{-1}$ and $1.5 \times 10^{44} \text{ erg s}^{-1}$ for \dot{E}_{bubble} and between $2 \times 10^{43} \text{ erg s}^{-1}$ and $2 \times 10^{45} \text{ erg s}^{-1}$ for \dot{E}_{AGN} , using single fiducial values (the average of the lower and upper limits of \dot{E} in a logarithmic scale) of $1.3 \times 10^{43} \text{ erg s}^{-1}$ and $2 \times 10^{44} \text{ erg s}^{-1}$ for \dot{E}_{bubble} and \dot{E}_{AGN} , respectively. The results of the two approaches indicate that the outflow in the [O III] cone is about 15 times more powerful than the outflow in the H α bubble, which reflects the fact that the former is faster and slightly more extended than the latter.

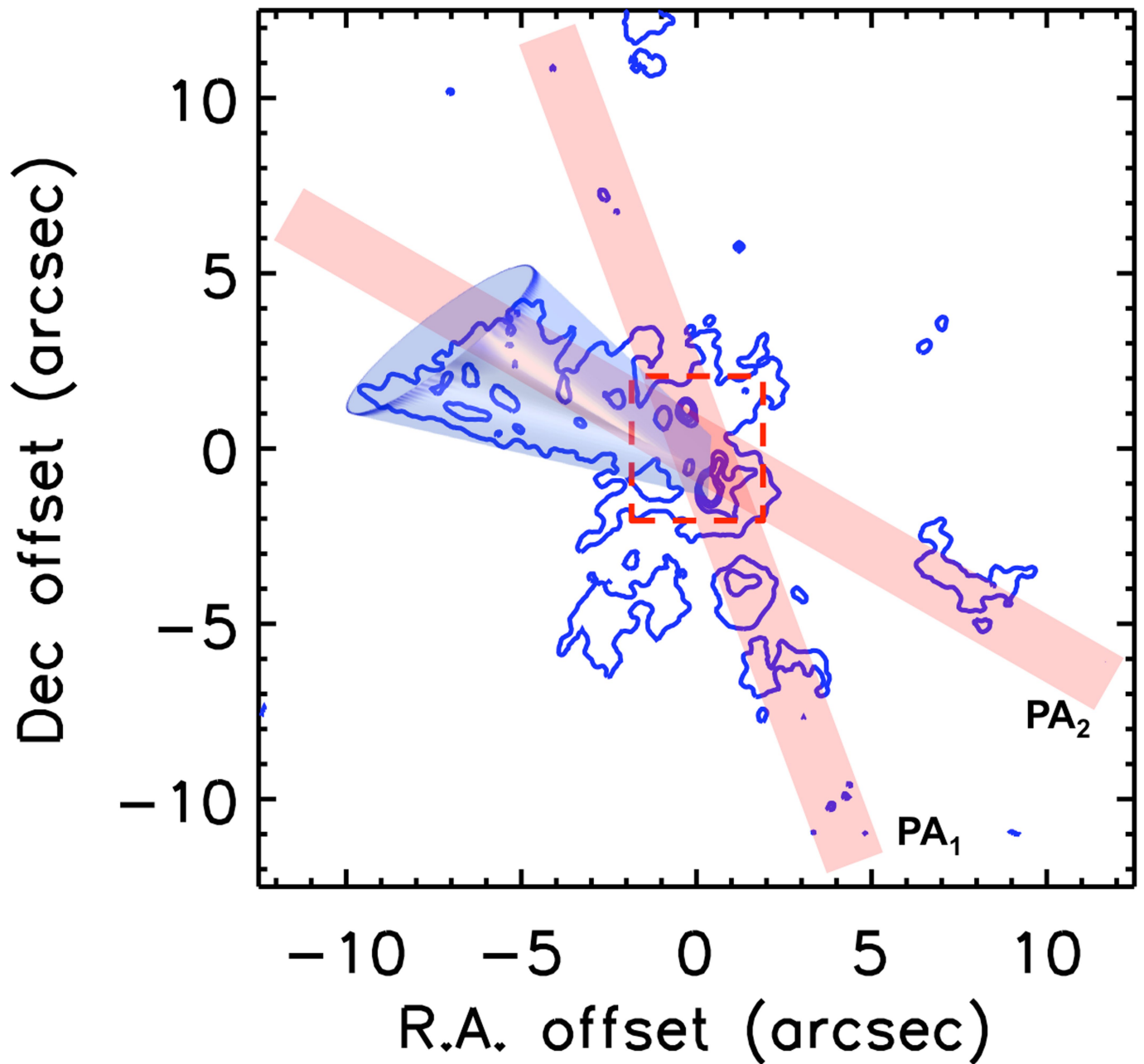
We can estimate the amount of mechanical energy returned from the nuclear starburst⁴ as $\dot{E}_{\text{mech}} = 7 \times 10^{41} \times [\text{SFR} (M_\odot \text{ yr}^{-1})] \text{ erg s}^{-1}$. Assuming an SFR of $100 M_\odot \text{ yr}^{-1}$ in the central region of the galaxy, the total injection of energy from the stars is $7 \times 10^{43} \text{ erg s}^{-1}$. Thus, star formation is consistent with powering the outflow in the H α bubble for the standard value of \dot{E}_{bubble} ($1.3 \times 10^{43} \text{ erg s}^{-1}$). In addition, energy injection from the stars could power the outflow in the [O III] cone at the lower limit of the energy range. However, we consider this unlikely, because this energy injection is below the fiducial value of kinetic power required to drive the outflow in the [O III] cone, $2 \times 10^{44} \text{ erg s}^{-1}$. The bolometric luminosity of the dual AGN estimated¹² from NuSTAR hard-X-ray data is $1.1_{-0.3}^{+0.5} \times 10^{45} \text{ erg s}^{-1}$, which is consistent with the value obtained¹³ by fitting the spectral energy distribution (about $2 \times 10^{45} \text{ erg s}^{-1}$). Our upper limit on the kinetic power of the [O III] cone is consistent with the bolometric luminosity of the dual AGN. We therefore conclude that the dual AGN is energetically capable of powering the outflow in the [O III] cone without the help of the starburst and that the mechanical energy injection rate from star formation is not powerful enough to accelerate the gas in this region.

In general, it is difficult to identify the driving mechanism of outflowing bubbles. In the case of the H α bubble, four pieces of evidence suggest that the outflow is driven by star formation. First, the nuclear starburst is energetically capable of driving the outflow without the need to invoke energy from an AGN. Second, the dynamical time of the outflow (about 7.4 Myr) is consistent with the age of the nuclear starburst in NGC 6240 (about $6\text{--}9 \text{ Myr}$)^{14,25,30}. Third, from a morphological point of view, a wind perpendicular to a nuclear disk is consistent with the structures of starburst-driven winds, in contrast to AGN-driven outflows, which have random orientations with respect to the galaxy disk³⁵. In NGC 6240, the position angle of the major axis of the H α bubble (about 110°) is almost perpendicular to the disk of the advanced merger (22°). On the other hand, the [O III] cone has a position angle of 56° , randomly oriented with respect to the galaxy disk. Finally, AGN-driven outflows usually have higher velocities than starburst-driven outflows^{4,35}. The [O III] cone exhibits gas clouds with velocity dispersion values of about $1,070 \text{ km s}^{-1}$, whereas the maximum dispersion of the Br γ emission is about 450 km s^{-1} (see also ref. ¹⁵).

Code availability. The SINFONI data used in this study were reduced with the public pipeline available at <https://www.eso.org/sci/software/pipelines/LINEFIT> and the routines used for reducing and analysing the APO/DIS long-slit data are available from the corresponding author upon request.

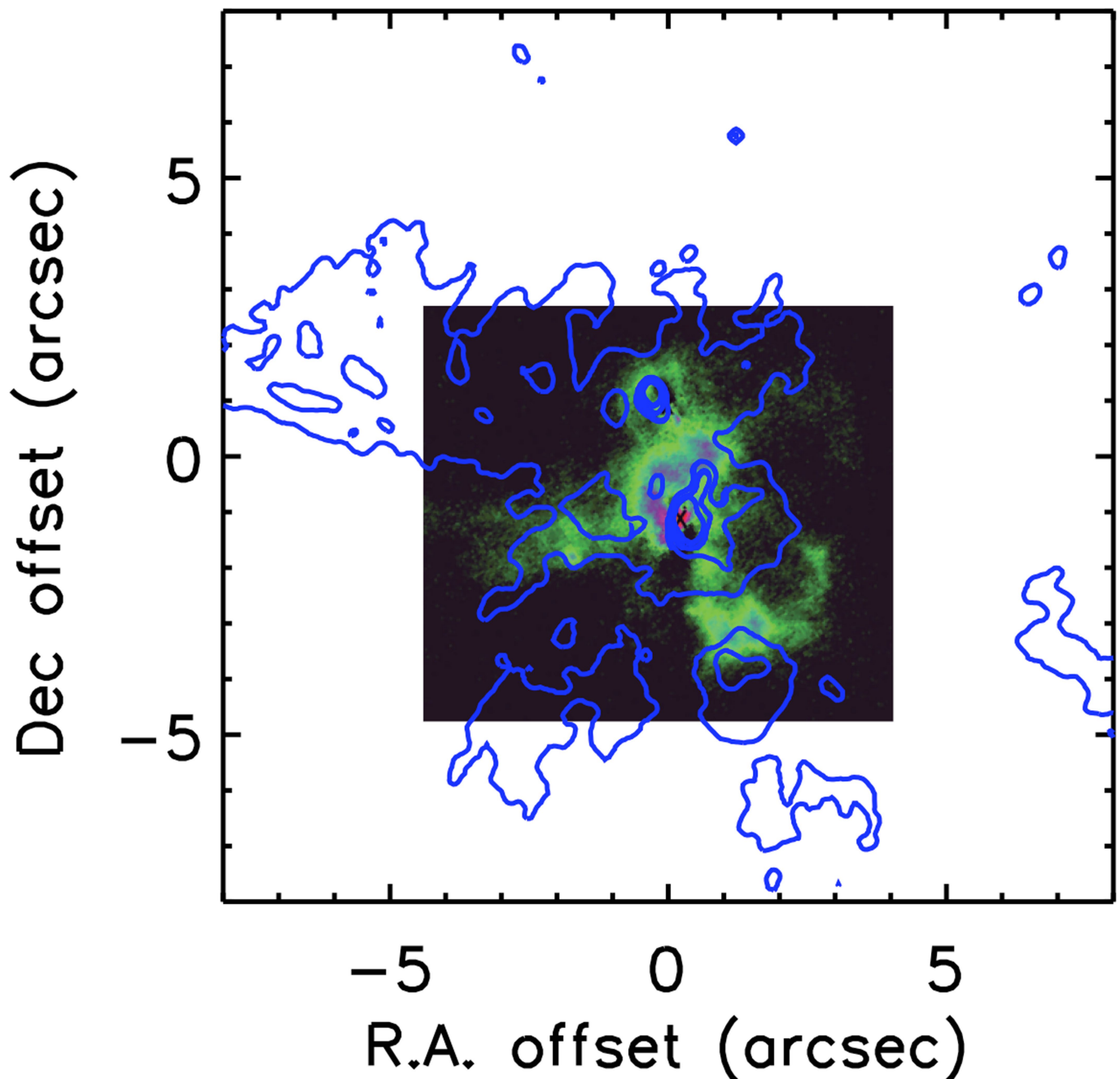
Data availability. The data plotted in the figures and that support other findings of this study are available from the corresponding author upon reasonable request. The SINFONI data used in this paper (programme 079.B-0576) can be obtained from the ESO Science Archive Facility (http://archive.eso.org/eso/eso_archive_main.html). The HST images (GO-12552) are available from MAST (<https://archive.stsci.edu/hst/>).

31. Eisenhauer, F. et al. in *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes* Vol. 4841 (eds Iye, M. & Moorwood, A. F. M.) 1548–1562 (International Society for Optics and Photonics, 2003).
32. Bonnet, H. et al. First light of SINFONI at the VLT. *Messenger* **117**, 17–24 (2004).
33. Davies, R. I. et al. How well can we measure the intrinsic velocity dispersion of distant disk galaxies? *Astrophys. J.* **741**, 69 (2011).
34. Dopita, M. A. & Sutherland, R. S. Spectral signatures of fast shocks. II. Optical diagnostic diagrams. *Astrophys. J.* **455**, 468–479 (1995).
35. Rupke, D. S. N. & Veilleux, S. The multiphase structure and power sources of galactic winds in major mergers. *Astrophys. J.* **768**, 75 (2013).
36. Medling, A. M. et al. Shocked gas in IRAS F17207-0014: ISM collisions and outflows. *Mon. Not. R. Astron. Soc.* **448**, 2301–2311 (2015).
37. Morwood, A. & Oliva, E. Infrared spectroscopy of forbidden [Fe II], H₂, and H line emission in galactic nuclei. *Astron. Astrophys.* **203**, 278–288 (1988).
38. Rodríguez-Ardila, A. et al. Molecular hydrogen and [Fe II] in active galactic nuclei. *Astron. Astrophys.* **425**, 457–474 (2004).
39. Riffel, R., Storchi-Bergmann, T. & Nagar, N. Near-infrared dust and line emission from the central region of Mrk1066: constraints from Gemini NIFS. *Mon. Not. R. Astron. Soc.* **404**, 166–179 (2010).
40. Puxley, P. J., Hawarden, T. G. & Mountain, C. M. Molecular and atomic hydrogen line emission from star-forming galaxies. *Astrophys. J.* **364**, 77–86 (1990).
41. van der Werf, P. et al. Near-infrared line imaging of NGC 6240 – collision shock and nuclear starburst. *Astrophys. J.* **405**, 522–537 (1993).
42. Max, C. et al. The core of NGC 6240 from Keck adaptive optics and Hubble space telescope NICMOS observations. *Astrophys. J.* **621**, 738–749 (2005).
43. Greene, J. E., Zakamska, N. L. & Smith, P. S. A spectacular outflow in an obscured quasar. *Astrophys. J.* **746**, 86–96 (2012).
44. Westmoquette, M. S., Smith, L. J. & Gallagher, J. S. III Spatially resolved optical integral field unit spectroscopy of the inner superwind of NGC 253. *Mon. Not. R. Astron. Soc.* **414**, 3719–3739 (2011).
45. Iono, D., Yun, M. S. & Mihos, J. C. Radial gas flows in colliding galaxies: connecting simulations and observations. *Astrophys. J.* **616**, 199–220 (2004).
46. Müller Sánchez, F. et al. Molecular gas streamers feeding and obscuring the active nucleus of NGC 1068. *Astrophys. J.* **691**, 749 (2009).
47. Schnorr-Müller, A. et al. Feeding and feedback in NGC 3081. *Mon. Not. R. Astron. Soc.* **457**, 972–985 (2016).
48. Steffen, W. et al. A 3D modeling tool for astrophysics. *IEEE Trans. Vis. Comput. Graph.* **17**, 454–465 (2011).
49. Cid Fernandes, R. et al. Alternative diagnostic diagrams and the ‘forgotten’ population of weak line galaxies in the SDSS. *Mon. Not. R. Astron. Soc.* **403**, 1036–1053 (2010).
50. Belfiore, F. et al. SDSS IV MaNGA – spatially resolved diagnostic diagrams: a proof that many galaxies are LIERs. *Mon. Not. R. Astron. Soc.* **461**, 3111–3134 (2016).



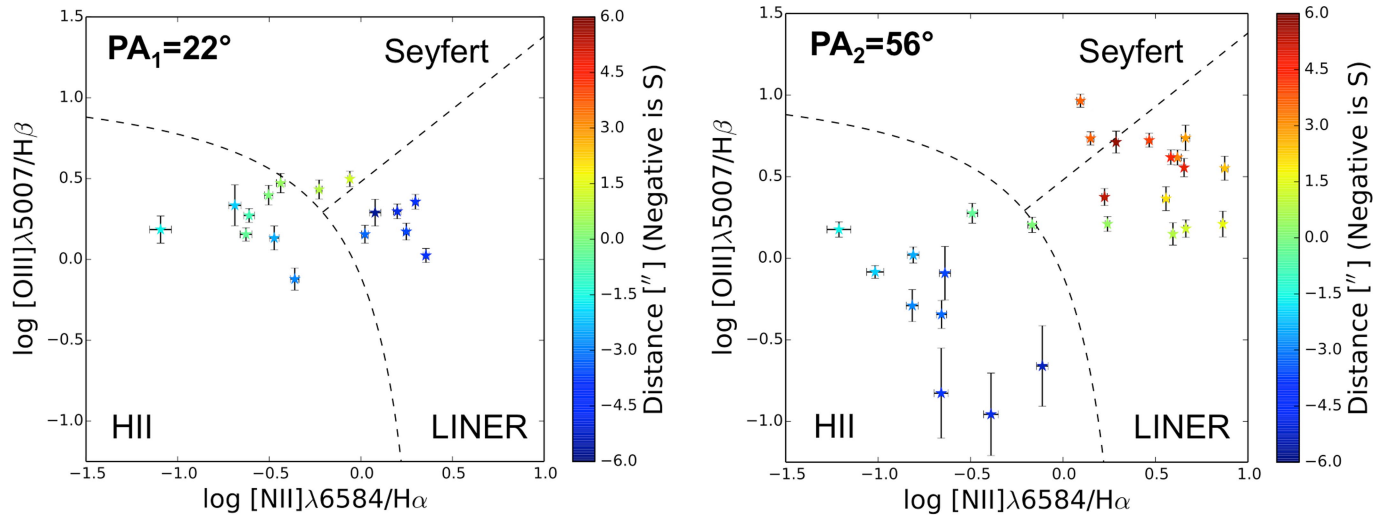
Extended Data Fig. 1 | Contour image of [O III] emission in NGC 6240. The blue curves show linear contours for the HST/F502N observations. The contours are set at 7.5%, 15%, 30%, 45%, 60%, 75% and 90% of the peak of emission. The extended [O III] emission is traced by the contour representing 7.5% of the peak of emission. The other contours (15–90% of the peak of emission) are located mostly around the two nuclei. A geometric model of the [O III] cone is shown in light blue. The model was created using the software Shape⁴⁸. We constrained the model (size and opening angle) to follow the outer contours (7.5% of the peak of emission) of the wedge-shaped structure in region 1. Interestingly, a regular cone (a cone with a sharp apex) does not provide a good fit to the wedge-shaped

structure. The best fit is obtained for a truncated cone. If we had used a regular cone, the apex would be located exactly at the position of the southwestern nucleus. This is consistent with our interpretation that the [O III] cone is probably produced by the two AGNs, with a larger contribution from the southwestern nucleus. For the [O III] cone, we obtained a size of 3.7 ± 0.2 kpc and an opening angle of $50.2 \pm 3.1^\circ$. The red-shaded rectangles indicate the spatial coverage of the long slits of the DIS. $PA_1 = 22^\circ$ is oriented along the major axis of the galaxy disk, and $PA_2 = 56^\circ$ covers the region where the [O III] cone is observed (region 1). Both slits were centred between the nuclei. The dashed rectangle represents the SINFONI field of view. North is up and east is to the left.



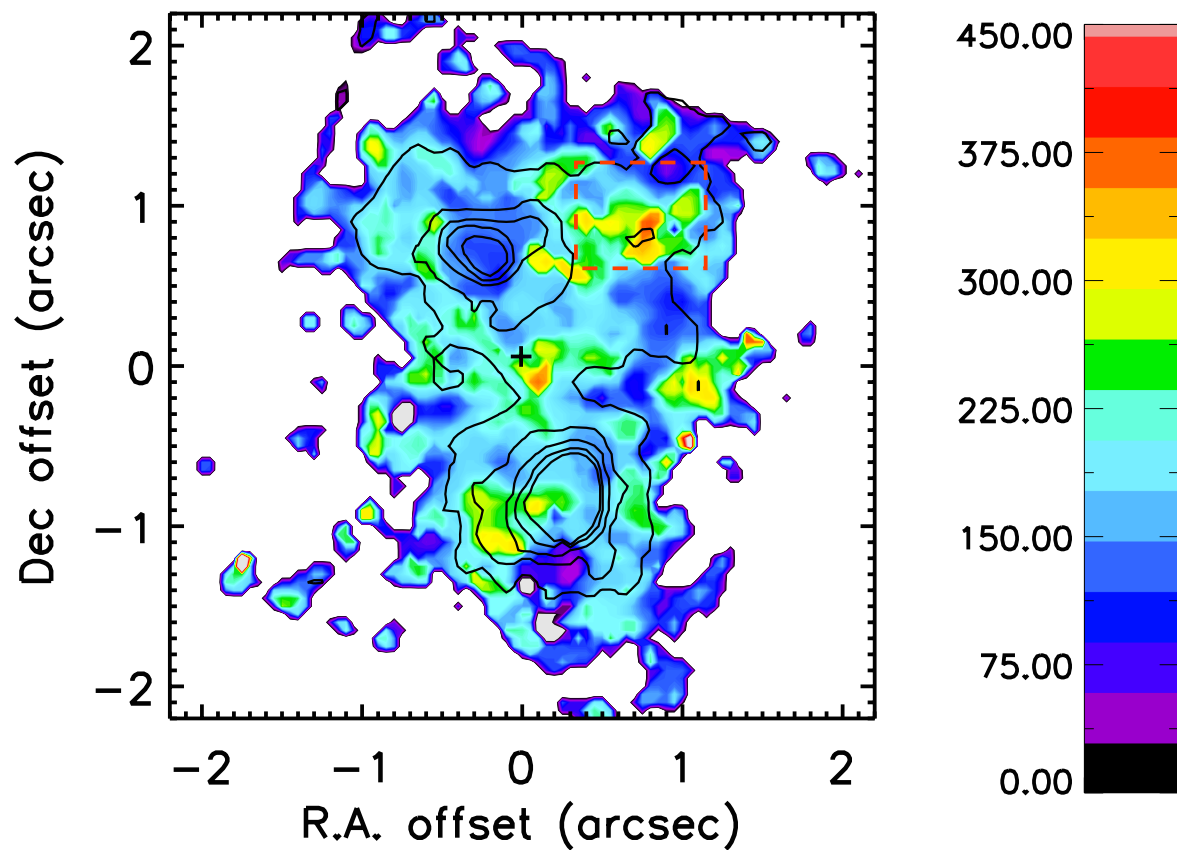
Extended Data Fig. 2 | Comparison of the morphologies of [O III] and H₂. An image of H₂ emission obtained with the near-infrared camera 2 (NIRC2)⁴² of the Keck adaptive optics system is superimposed on the [O III] contours from Extended Data Fig. 1. Black represents fluxes < 0.011 of the peak of emission. The absence of molecular gas at the locations of the [O III] cone and the H α bubble clearly indicates that these two

structures are located in regions that are not greatly influenced by the merger process. By contrast, the majority of perturbations caused by the merger activity are seen in the central region between the nuclei, and as gas streamers in the regions east and southwest of the SW nucleus (regions 3 and 4 in our analysis; see also Fig. 2b).



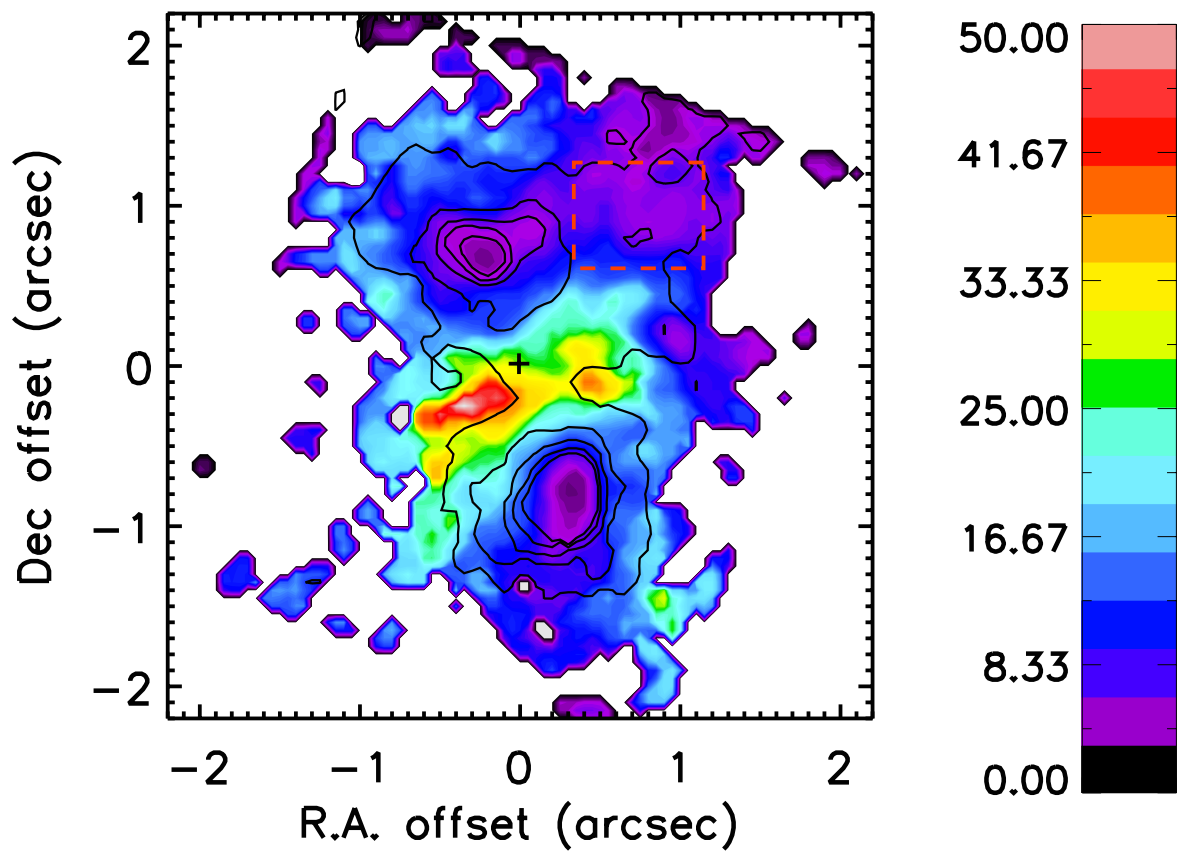
Extended Data Fig. 3 | Optical emission-line diagnostic diagrams. The galaxy was observed at two position angles, $PA_1 = 22^\circ$ and $PA_2 = 56^\circ$ (see Extended Data Fig. 1). The positive values of angular distance (green to red in the colour bar) correspond to the direction north of the centre of the galaxy, at that position angle. Negative angular distance values (green to blue in the colour bar) correspond to the direction south of the centre of the galaxy, at that position angle. The BPT diagram is usually divided into three regions: AGN (or Seyfert), LINER (or LIER, low-ionization emission-line region; see also ref. ⁵⁰) and H II (or starburst region). In both

panels, we plot the extreme starburst diagnostic line²¹ (curved dashed line) and the LIER/LINER diagnostic line⁴⁹ (straight dashed line). H β emission was detected with a signal-to-noise ratio higher than 3 in 16 spatial elements at $PA_1 = 22^\circ$ (from $r = -6''$ to $r = 2''$) and in 26 spatial elements at $PA_2 = 56^\circ$ (from $r = -5''$ to $r = 6''$). There are 15 spatial elements inside the [O III] cone (Fig. 3). The error bars correspond to the uncertainties of the flux ratios (one standard deviation) and were calculated via standard error propagation for the flux of each emission line.



Extended Data Fig. 4 | Map of Br γ velocity dispersion. The contours delineate the Br γ flux distribution and are set at 15%, 30%, 45%, 60%, 75% and 90% of the peak of emission. The dashed rectangle delimits the base of the H α bubble. Regions in white correspond to pixels where the Br γ flux

is less than 5% of the peak of emission and thus were masked out. North is up and east is to the left. The colour bar indicates the range of velocity dispersion values observed in units of kilometres per second.



Extended Data Fig. 5 | Map of $H_2/Br\gamma$ flux ratio. The contours delineate the $Br\gamma$ flux distribution and are set at 15%, 30%, 45%, 60%, 75% and 90% of the peak of emission. The dashed rectangle delimits the base of the $H\alpha$

bubble. Regions in white correspond to pixels where the $Br\gamma$ flux is less than 5% of the peak of emission and thus were masked out. North is up and east is to the left. The colour bar indicates the range of ratios observed.

Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip

Amir H. Atabaki^{1,11*}, Sajjad Moazeni^{2,11}, Fabio Pavanello^{3,6,11}, Hayk Gevorgyan⁴, Jelena Notaros^{3,7}, Luca Alloatti^{1,8}, Mark T. Wade^{3,9}, Chen Sun^{2,9}, Seth A. Kruger⁵, Huaiyu Meng¹, Kenaish Al Qubaisi⁴, Imbert Wang⁴, Bohan Zhang⁴, Anatol Khilo⁴, Christopher V. Baiocco⁵, Miloš A. Popović^{3,4}, Vladimir M. Stojanović² & Rajeev J. Ram¹

Electronic and photonic technologies have transformed our lives—from computing and mobile devices, to information technology and the internet. Our future demands in these fields require innovation in each technology separately, but also depend on our ability to harness their complementary physics through integrated solutions^{1,2}. This goal is hindered by the fact that most silicon nanotechnologies—which enable our processors, computer memory, communications chips and image sensors—rely on bulk silicon substrates, a cost-effective solution with an abundant supply chain, but with substantial limitations for the integration of photonic functions. Here we introduce photonics into bulk silicon complementary metal–oxide–semiconductor (CMOS) chips using a layer of polycrystalline silicon deposited on silicon oxide (glass) islands fabricated alongside transistors. We use this single deposited layer to realize optical waveguides and resonators, high-speed optical modulators and sensitive avalanche photodetectors. We integrated this photonic platform with a 65-nanometre-transistor bulk CMOS process technology inside a 300-millimetre-diameter-wafer microelectronics foundry. We then implemented integrated high-speed optical transceivers in this platform that operate at ten gigabits per second, composed of millions of transistors, and arrayed on a single optical bus for wavelength division multiplexing, to address the demand for high-bandwidth optical interconnects in data centres and high-performance computing^{3,4}. By decoupling the formation of photonic devices from that of transistors, this integration approach can achieve many of the goals of multi-chip solutions⁵, but with the performance, complexity and scalability of ‘systems on a chip’^{1,6–8}. As transistors smaller than ten nanometres across become commercially available⁹, and as new nanotechnologies emerge^{10,11}, this approach could provide a way to integrate photonics with state-of-the-art nanoelectronics.

Sustained innovations in electronics, predominantly in CMOS, have transformed computing, communications, sensing and imaging. More recently, silicon photonics has been leveraging the CMOS infrastructure to address the growing demands for optical communications for internet and data centre networks^{3,6,7}. This convergence of photonics with CMOS promises to transform electronic–photonic technologies, enabling processor and memory chips with high-bandwidth optical input/output^{1,4}, communications chips with high-fidelity optical signal processing^{2,12}, and highly parallel optical biochemical sensors for blood analysis¹³ and gene sequencing¹⁴. To make these a reality, photonic devices need to be integrated with a variety of nanoelectronic functions (digital, analogue, memory, storage and so on) on a single silicon die (chip).

Monolithic (that is on a single chip) integration of photonic devices in close proximity to electronic circuits is crucial for two main reasons: it allows us to achieve the required levels of performance, scalability

and complexity simultaneously for electronic–photonic systems; and substantially accelerates system-level innovation by enabling a cohesive design environment and device ecosystem to realize entire ‘systems on a chip’. In fact, the accelerated progress in recent years in electronics is a direct result of such a system-on-a-chip approach and the addition of new functions and components to CMOS to create new monolithic device platforms, such as wireless communications and radar imaging chips (through the addition of inductors and transmission lines¹⁵) and image sensors (through silicon photodiodes¹⁶).

The greatest challenge towards the integration of photonic circuits into CMOS has been the lack of a semiconductor material with suitable optical properties for realizing active and passive photonic functions in bulk CMOS, which is the dominant manufacturing platform for micro-electronic chips (every Intel, Apple and Nvidia CPU/GPU, all computer memory and flash storage, and so on). As a result, all efforts so far to integrate photonics into CMOS have been limited to silicon-on-insulator (SOI) substrates^{1,6–8}. These processes are cost-prohibitive for many applications (for example, computer memory) and have a limited supply chain for high volume markets. The same photonic integration challenge also exists for the leading CMOS technologies below 28-nm transistor nodes—fin field effect transistor (FinFET) and thin-body fully depleted SOI¹⁷ (TBF-D-SOI)—where the crystalline silicon layers are too thin (less than 20 nm) to support photonic structures with sufficient optical confinement. To address these integration challenges, we have developed a photonic platform using an optimized polycrystalline silicon (polysilicon) film that could be deposited on silicon oxide islands that are ubiquitous in CMOS (used to isolate transistors) even in the most recent technologies using FinFET and TBF-D-SOI¹⁷ (Fig. 1a).

Deposited electronic and photonic devices on glass have already affected many fields: thin-film transistors have enabled today’s display technologies, and photonic platforms with thin-film components on glass have been commercially deployed in optical communications systems¹⁸. However, deposited photonic components have been restricted to passive functions (for example, filters and delay lines) lacking light detection and modulation. A variety of materials, including amorphous and polycrystalline silicon^{19–21}, polymer-based devices²² and chalcogenides²³, have been deposited on glass in the attempt to realize active photonic components. Nevertheless, the integration of a fully functional photonic platform (that is, passive functions, optical modulators and detectors) and its integration with CMOS nanoelectronics is yet to be demonstrated. In this work, we have integrated a fully functional polysilicon photonic platform with a 65-nm bulk CMOS process through the addition of a few extra processing steps without affecting the transistors’ native performance, and demonstrated large-scale monolithic electronic–photonic systems.

Figure 1a shows transistor structures in today’s three dominant deeply scaled CMOS processes. The silicon oxide shallow trench

¹Massachusetts Institute of Technology, Cambridge, MA, USA. ²University of California, Berkeley, Berkeley, CA, USA. ³University of Colorado, Boulder, Boulder, CO, USA. ⁴Boston University, Boston, MA, USA. ⁵Colleges of Nanoscale Science and Engineering, State University of New York (SUNY) Polytechnic Institute, Albany, NY, USA. ⁶Colleges of Nanoscale Science and Engineering, State University of New York (SUNY) Polytechnic Institute, Albany, NY, USA. ⁷Present address: Photonics Research Group, Ghent University-IMEC and Center for Nano- and Bio-photonics, Ghent, Belgium. ⁸Present address: Massachusetts Institute of Technology, Cambridge, MA, USA. ⁹Present address: Institute of Electromagnetic Fields (IEF), ETH Zurich, Zurich, Switzerland. ¹⁰Present address: Ayar Labs, Inc., Emeryville, CA, USA. ¹¹These authors contributed equally: Amir H. Atabaki, Sajjad Moazeni, Fabio Pavanello. *e-mail: atabaki@mit.edu

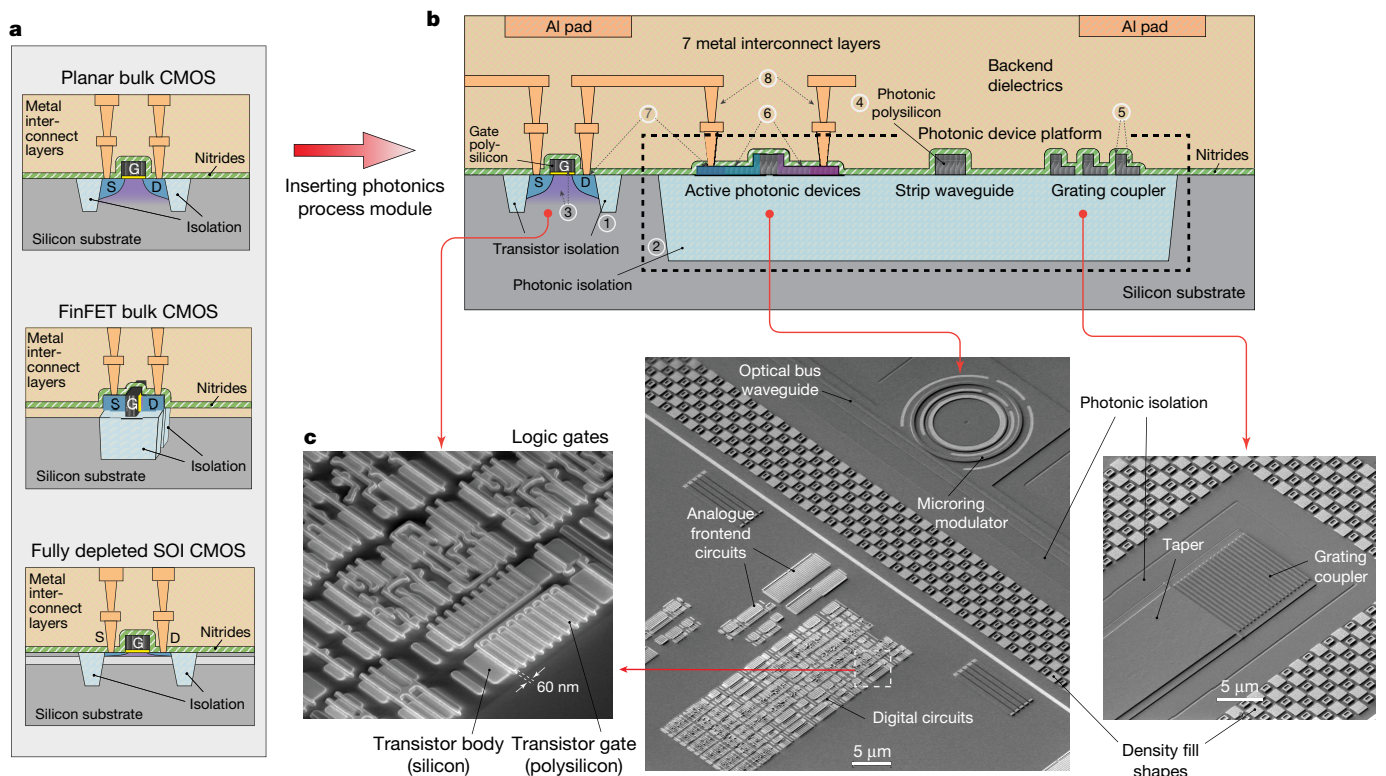


Fig. 1 | Photonic integration with nanoscale transistors. **a**, Illustration of three major deeply scaled CMOS processes: planar bulk CMOS, FinFET bulk CMOS, and fully depleted SOI CMOS. **b**, Integration of a photonics process module into planar bulk CMOS with photonic devices implemented in an optimized polysilicon film (220 nm) deposited on a photonic trench filled with silicon oxide (about 1.5 μm). The numbers indicate major fabrication steps in the order appearing in the process: (1) and (2), transistor and photonic isolation fabrication; (3) transistor

frontend fabrication up to source/drain implant, including gate definition; (4) deposition, annealing and polishing of photonic polysilicon film; (5) polysilicon full and partial etching for forming strip and ridge photonic structures; (6) doping implants (P and N) for active photonics; (7) high doping implants (P++ and N++) and salicidation for both electronic and photonic devices; and (8) metallization. **c**, Scanning electron micrographs of different photonic and electronic blocks in our monolithic platform.

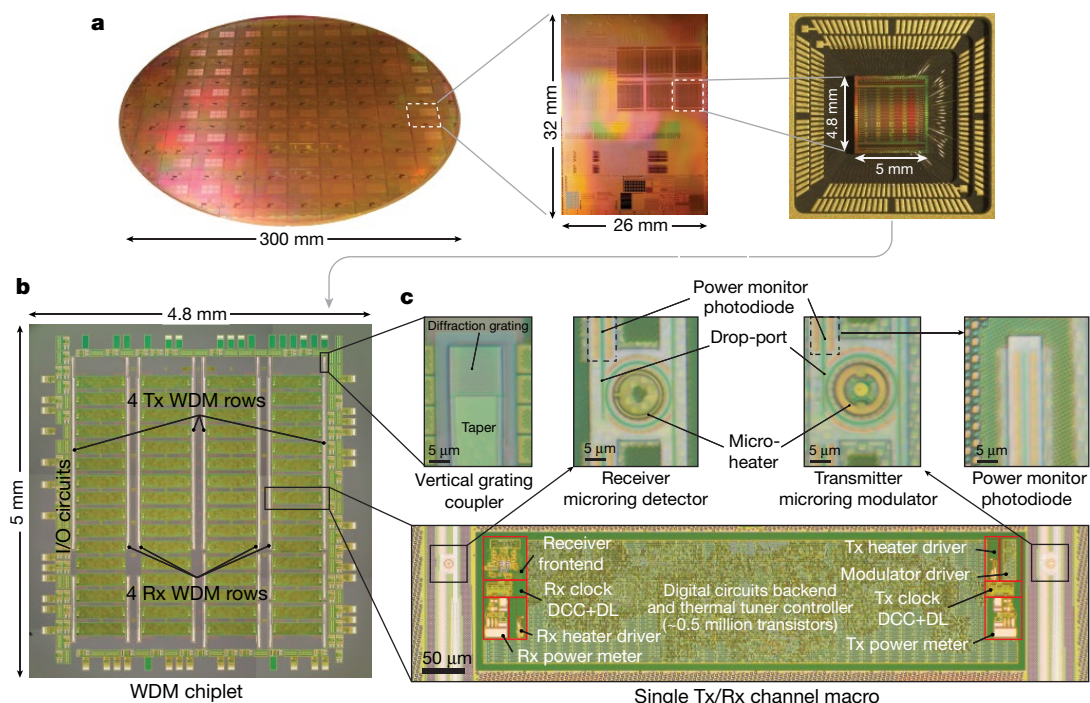


Fig. 2 | Monolithic electronic-photonics platform in 65-nm bulk CMOS. **a**, Photograph of a fully fabricated 300-mm wafer with monolithic electronics and photonics, and close-ups of a reticle on this wafer, and a packaged WDM chiplet. **b**, Micrograph of a WDM chiplet with four

transmitter (Tx) and receiver (Rx) rows. **c**, Close-up of a single transceiver macro and its photonic and electronic circuit components, such as the grating coupler, optical modulator and detector, and power monitor photodiodes. I/O, input/output; DCC, duty-cycle corrector; DL, delay line.

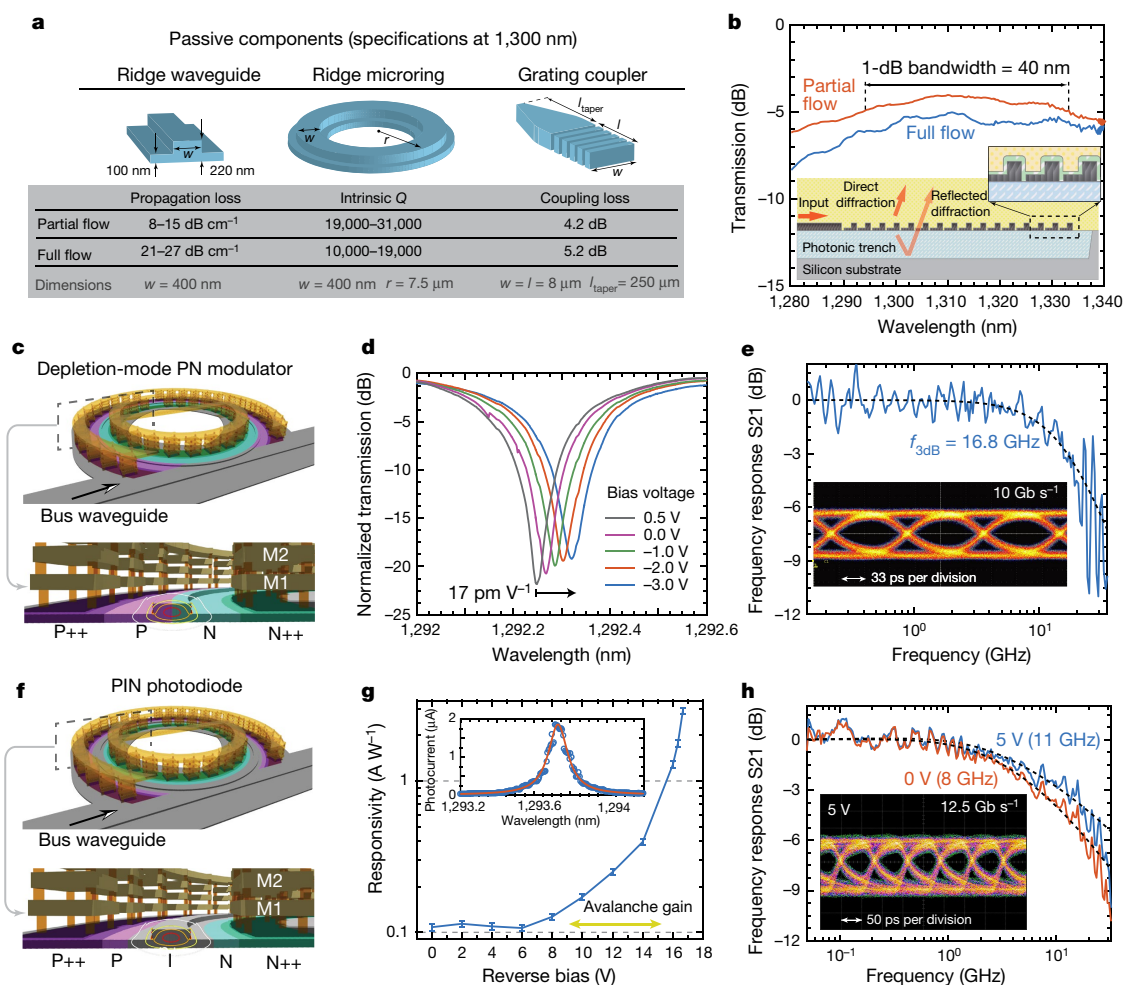


Fig. 3 | Photonics platform performance. **a**, Passive component specifications at 1,300 nm for partial- and full-flow wafers. **b**, Transmission spectrum and the longitudinal cross-section of the grating coupler (inset). **c**, Microring modulator three-dimensional layout. **d**, Transmission spectrum of a modulator resonance with loaded Q -factor of 5,000. **e**, Modulator electro-optic frequency response (S21) and the eye-diagram obtained with 2 V peak-to-peak drive voltage (V_{pp}), extinction ratio of 5.7 dB and insertion loss of 3.8 dB. **f**, Microring photodiode three-

dimensional layout. M1 and M2 are the first and second metal layers in the process. P++ and N++ are the high doping regions under metal contacts and I is the intrinsic region in the photodiode. **g**, Responsivity versus reverse bias voltage. Avalanche gain is observed at biases above 8 V. **h**, Photodiode frequency response (S21) under 0 V and 5 V reverse bias with 3-dB bandwidths of 8 GHz and 11 GHz, respectively. The inset shows the eye diagram obtained under 5 V bias.

isolation for transistors in advanced CMOS nodes is too thin to support low-loss optical waveguides on top of this layer owing to light leakage into the substrate. We address this issue by locally adding a thicker silicon oxide photonic isolation layer (about 1.5 μ m) with a fabrication process very similar to shallow trench isolation. An optimized polysilicon film (220 nm thick) with low optical propagation loss and high carrier mobility is then deposited on this layer, and is used for passive photonic components, free-carrier plasma dispersion modulators^{24,25}, and photodetectors that make use of the absorption by defect states at polysilicon grain boundaries^{26,27}. Photonic isolation layer fabrication and polysilicon film deposition are followed by two etching steps (full and partial, for strip and ridge structures) and two doping implants (N-type and P-type for modulators and detectors) to form our photonics process module that is inserted into the CMOS fabrication process flow. Figure 1b shows the cross-sectional drawing of three representative photonic components in our polysilicon photonic platform next to a transistor in a planar bulk CMOS process.

The photonics process module is inserted in the middle of transistor processing, after gate definition, but before source and drain implants (see numbers in Fig. 1b for the fabrication order). With this approach, all of the high-temperature photonics processing takes place before the definition of the source, drain and channel of transistors. This eliminates the need for re-optimizing the source and drain implants

and anneal processes that would otherwise be needed because of the sensitivity of deeply scaled transistors to the source and drain doping profiles²⁸. Also, this approach allows us to reuse some of the frontend processing steps (high-doping implants, and silicide formation) for active photonic components to minimize the number of photolithography masks. In doing so, the entire fabrication development is shifted to the photonics side, because low-loss photonic structures have to be implemented while transistor gate features already exist on the same level. This necessitates careful optimization of the polysilicon film deposition, polishing, and etching steps to achieve low (< 1 nm) surface and sidewall roughness for low-loss and high-performance devices (see Methods for process details).

This optimized photonics process was integrated with an entire commercial 65-nm bulk CMOS process with seven metal interconnect layers, featuring transistors with three different threshold voltages and two oxide thickness variants. Figure 1c shows bird's eye scanning electron micrographs of our monolithic platform with photonic components next to transistors with 60 nm channel length. This platform is fabricated on 300-mm-diameter wafers (the largest size in production at present) in a CMOS foundry located at the Colleges for Nanoscale Sciences and Engineering, SUNY Polytechnic Institute, Albany, New York. Figure 2a shows a photo of a fully fabricated wafer, and close-up photos of the entire reticle and one packaged chiplet composed of

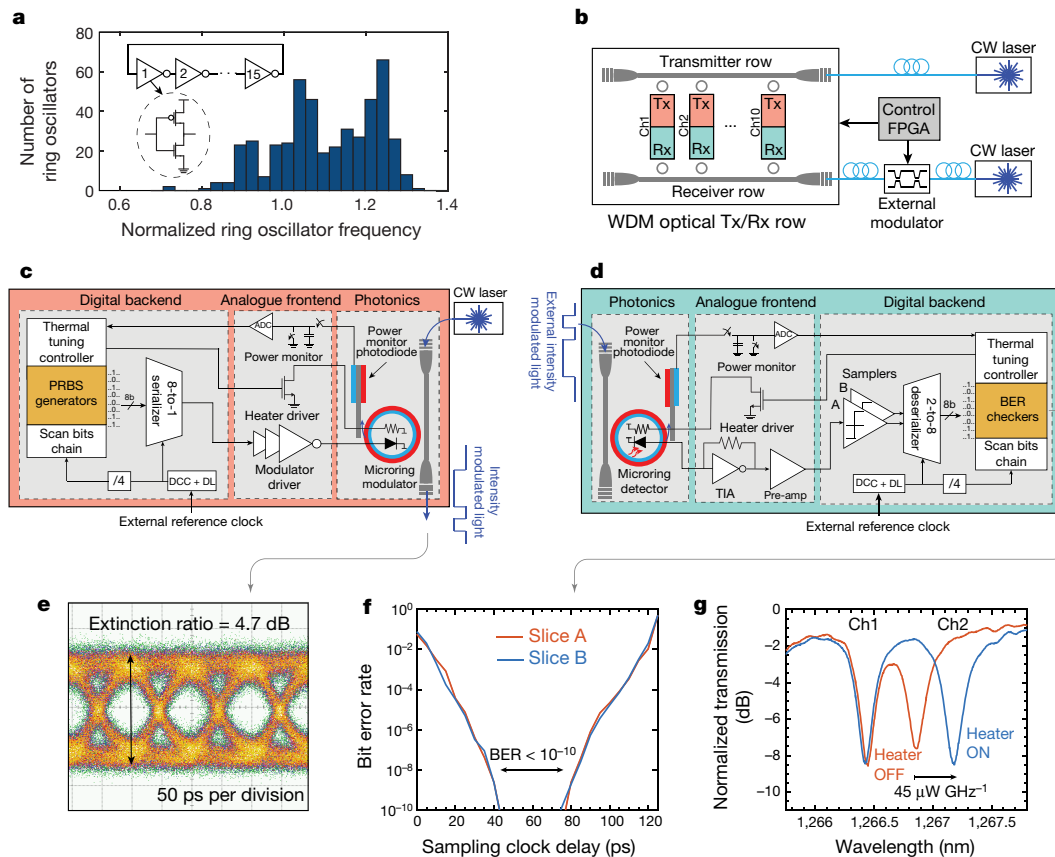


Fig. 4 | Electro-optical testing of WDM transceiver chips. **a**, Histogram of measured frequencies of 485 test ring oscillators normalized to the frequency obtained from simulation with the native CMOS process design kit models. FPGA, field-programmable gate array. **b**, Block diagram of one WDM transmit-receive row in our test setup. **c**, Block diagram of one transmitter channel. **d**, Block diagram of one receiver channel.

TIA, transimpedance amplifier. **e**, 10 Gb s⁻¹ transmit eye diagram using the on-chip pseudorandom binary sequence (PRBS) generator. **f**, 7 Gb s⁻¹ receiver bathtub curve obtained from the on-chip bit-error-rate (BER) tester by sweeping the delay between the clocks for the receiver and external transmitter (sampling clock delay). **g**, Thermal tuning of one WDM channel using the integrated microheater.

several wavelength division multiplexing (WDM) photonic transceivers. The micrographs of the transceiver chiplet, transmitter and receiver circuit blocks, and individual photonic components are shown in Fig. 2b and c. We were able to build a library of passive and active photonic components (waveguides, microring resonators, vertical grating couplers, high-speed modulators, and avalanche photodetectors) with a performance similar or better than previous demonstrations on polysilicon^{20,21,26}, next to circuit blocks composed of millions of transistors operating at native CMOS process specifications.

Figure 3a summarizes the performance of passive components measured on partial-flow (passive photonics only) and full-flow (active and passive photonics with electronics) wafers, at a wavelength of 1,300 nm. We achieved a propagation loss of approximately 10 dB cm⁻¹ for ridge and strip waveguides, and a loaded quality factor (Q-factor) of >20,000 for microring resonators on partial-flow wafers. Full-flow wafers exhibit higher loss, but this issue did not have much effect on the performance of our optical transceivers: the 20 dB cm⁻¹ waveguide loss results in a loss of 3 dB across the 10-lambda (that is, wavelength) WDM rows, and the loaded Q-factor of 10,000 of microring resonators is close to optimal for resonant modulators and detectors of bandwidth 10–20 GHz (see Methods for further discussion). Waveguide loss and resonator Q-factor are two times better at 1,550 nm (Extended Data Fig. 1), but all optical transceivers were initially designed at 1,300 nm. Grating couplers for coupling light into and out of the chip are designed using both the partial- and full-etch steps to construct a periodic L-shaped geometry (Fig. 3b). The measured grating transmission, shown in Fig. 3b, indicates a peak efficiency of -4.2 dB for the partial flow and -5.2 dB for the full flow, with 1-dB bandwidth of around 40 nm (see Methods for discussions of further device improvements).

Depletion-mode resonant modulators and defect-based photodetectors were implemented using ridge microring structures with lateral PN and PIN diode junctions, respectively (where P, N and I refer to P-type, N-type and Intrinsic). Two mid-level doping implants, P-type and N-type with concentrations of $6 \times 10^{18} \text{ cm}^{-3}$, are optimized for these devices. Device micrographs and designs are shown in Figs. 2c and 3c and f (see Methods for discussion on implants).

The modulator uses a lateral PN junction to modulate light through the modulation of the resonance wavelength using the free-carrier plasma dispersion effect²⁹ (Fig. 3d). By operating the PN junction under full depletion, a 3-dB bandwidth of 16.8 GHz (Fig. 3e) and digital modulation at 10 Gb s⁻¹ is achieved with only 2 V V_{pp} modulation signal (inset to Fig. 3e).

The defect-based photodetector has a responsivity of 0.11 A W⁻¹ (quantum efficiency of 10%) near 1,300 nm under very low bias voltages (Fig. 3g) using a resonant design that enhances the weak absorption in polysilicon (inset to Fig. 3g). We also observed avalanche gain for the first time in polysilicon photodetectors³⁰ at bias voltages above 8 V (Fig. 3g), leading to a responsivity of 1.3 A W⁻¹ at bias 16 V with a noise equivalent power of 0.27 pW Hz^{-1/2}. This device has a 3-dB bandwidth of more than 8 GHz under reverse bias voltages above 0 V, reaching 11 GHz for a bias of 5 V (Fig. 3h). More results on photodetectors are given in Extended Data Fig. 2.

To examine the performance of transistors after introducing the photonics module into the CMOS process, electrical ring oscillators composed of 15 equally sized inverting stages were used inside all electronic-photonics blocks to probe the speed of transistors, as well as the intra- and inter-die variations. The fastest transistors (low threshold voltage) in the process with gate lengths of 55 nm were used for the ring

oscillator design. Figure 4a shows the histogram of the normalized frequency of ring oscillators relative to the nominal frequency (2.33 GHz, single stage delay of 14.3 ps), simulated with the original process design kit provided by the foundry. This distribution is within the standard range of native CMOS processes, confirming that our photonic process module does not degrade the performance of transistors.

As a first demonstration of monolithic electronic–photonic systems in this platform, we have implemented high-bandwidth photonic WDM transceivers. We designed a total of six chiplets, each containing four stand-alone WDM transmitter and receiver rows, each supporting up to 16 channels. Different designs for resonant modulators and detectors were used in WDM rows. The chiplets are diced and wire-bond packaged with 100 pads to provide direct-current (d.c.) supplies, bias signals and high-speed clocks for electro-optical testing (Fig. 2a). By integrating all of the analogue and digital blocks, signal generation and error estimation of transceivers can be performed on the chip.

The transmitter is composed of a full digital backend that generates a pseudorandom binary sequence signal, a serializer with an 8 to 1 ratio, and finally an inverter chain that drives the microring modulator (Fig. 4c). On the receiver side, a transimpedance amplifier analogue frontend converts and amplifies the received photo-current into a voltage signal, and a pair of double-data-rate samplers converts the signal into the digital domain³¹ (Fig. 4d). These bits are deserialized and fed into a bit-error-rate checker on the chip. The generated pseudorandom binary sequence signal and bit-error-rate data are monitored via on-chip scan chains to measure the functionality and performance of the transceivers. Overall, approximately 30,000 logic gates (about 0.5 million transistors) from digital standard cells have been used in each transceiver channel.

The operation of one channel of a 10-lambda WDM transceiver is shown in Fig. 4e and f. Modulators were operated in the depletion mode (voltage swing from 0 to -1.5 V) at 10 Gb s^{-1} data-rate with an extinction ratio of 4.7 dB (Fig. 4e). The receiver achieved a bit-error rate better than 10^{-10} with -3 dBm input optical power at 7 Gb s^{-1} , as shown in Fig. 4f. The speed of receivers is limited at present by the long on-chip clock distribution network and could be further improved by integrating local clock generators³¹. Thermal tuning controllers and heater drivers are also included in the transceivers, to adjust for microring resonance fluctuations due to temperature and process variations³². Using the digital-to-analogue converters in the thermal tuning controllers and heater drivers, we measured a tuning efficiency of $45\mu\text{W GHz}^{-1}$ for integrated microheaters on microring modulators and detectors (Fig. 4g). The total electrical energy consumption of the transmitter and receiver including the serializer and deserializer was 100 fJ b^{-1} and 500 fJ b^{-1} , respectively. The transceiver achieved a bandwidth density of $180\text{ Gb s}^{-1}\text{ mm}^{-2}$ with 10% of the effective area occupied by photonics, which can be reduced to 5% by optimizing the floorplan. Incorporating this photonics platform in advanced sub-10-nm technology nodes with higher transistor densities³³ would lead to $>2\text{ Tb s}^{-1}\text{ mm}^{-2}$ bandwidth densities meeting the needs of next-generation systems on a chip.

The optical transceivers in bulk CMOS demonstrated here are an important milestone towards multi-terabytes-per-second optical interconnects for direct integration with logic and memory to improve the performance of computing systems, at present limited by the chip input/output bandwidth. This photonic platform and integration approach illustrates how adding photonic functions onto a variety of substrates could enable the next generation of systems on a chip for computing, communications, imaging and sensing.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0028-z>

Received: 5 October 2017; Accepted: 19 February 2018;
Published online 18 April 2018.

- Sun, C. et al. Single-chip microprocessor that communicates directly using light. *Nature* **528**, 534–538 (2015).
- Khilo, A. et al. Photonic ADC: overcoming the bottleneck of electronic jitter. *Opt. Express* **20**, 4454–4469 (2012).
- Rumley, S. et al. Silicon photonics for exascale systems. *J. Lightwave Technol.* **33**, 547–562 (2015).
- Miller, D. A. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–749 (2000).
- Boeuf, F. et al. A multi-wavelength 3D-compatible silicon photonics platform on 300 mm SOI wafers for 25 Gb/s applications. In *Electron Devices Meeting (IEDM)* 13–3 (IEEE, 2013).
- Assefa, S. et al. Monolithic integration of silicon nanophotonics with CMOS. In *Photonics Conference (IPC)* 626–627 (IEEE, 2012).
- Narasimha, A. et al. A 40-Gb/s QSPF optoelectronic transceiver in a $0.13\mu\text{m}$ CMOS silicon-on-insulator technology. In *Optical Fiber Communication Conference (OMK7)* (Optical Society of America, 2008).
- Awmy, A. et al. A 40 Gb/s monolithically integrated linear photonic receiver in a $0.25\mu\text{m}$ BiCMOS SiGe:C technology. *IEEE Microw. Wirel. Compon. Lett.* **25**, 469–471 (2015).
- leong, M., Doris, B., Kedzierski, J., Rim, K. & Yang, M. Silicon device scaling to the sub-10-nm regime. *Science* **306**, 2057–2060 (2004).
- del Alamo, J. A. Nanometre-scale electronics with III–V compound semiconductors. *Nature* **479**, 317–323 (2011).
- Desai, S. B. et al. MoS_2 transistors with 1-nanometer gate lengths. *Science* **354**, 99–102 (2016).
- Ghelfi, P. et al. A fully photonics-based coherent radar system. *Nature* **507**, 341–345 (2014).
- Mudumba, S. et al. Photonic ring resonance is a versatile platform for performing multiplex immunoassays in real time. *J. Immunol. Methods* **448**, 34–43 (2017).
- Quail, M. A. et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
- Zolfaghari, A., Chan, A. & Razavi, B. Stacked inductors and transformers in CMOS technology. *IEEE J. Solid-State Circuits* **36**, 620–628 (2001).
- Fossum, E. R. CMOS image sensors: electronic camera-on-a-chip. *IEEE Trans. Electron Dev.* **44**, 1689–1698 (1997).
- Magarshack, P., Flatresse, P. & Cesana, G. UTBB FD-SOI: A process/design symbiosis for breakthrough energy-efficiency. In *Proc. Conf. Design, Automation and Test in Europe* 952–957 (EDA Consortium, 2013).
- Wörhoff, K., Heideman, R. G., Leinse, A. & Hoekman, M. TriPleX: a versatile dielectric photonic platform. *Adv. Opt. Technol.* **4**, 189–207 (2015).
- Selvaraja, S. K. et al. Low-loss amorphous silicon-on-insulator technology for photonic integrated circuitry. *Opt. Commun.* **282**, 1767–1770 (2009).
- Sun, C. et al. A monolithically-integrated chip-to-chip optical link in bulk CMOS. *IEEE J. Solid-State Circ.* **50**, 828–844 (2015).
- Preston, K., Manipatruni, S., Gondarenko, A., Poitras, C. B. & Lipson, M. Deposited silicon high-speed integrated electro-optic modulator. *Opt. Express* **17**, 5118–5124 (2009).
- Dangel, R. et al. Polymer-waveguide-based board-level optical interconnect technology for datacom applications. *IEEE Trans. Adv. Packag.* **31**, 759–767 (2008).
- Eggleton, B. J., Luther-Davies, B. & Richardson, K. Chalcogenide photonics. *Nat. Photon.* **5**, 141–148 (2011).
- Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005).
- Shainline, J. M. et al. Depletion-mode carrier-plasma optical modulator in zero-change advanced CMOS. *Opt. Lett.* **38**, 2657–2659 (2013).
- Mehta, K. K. et al. Polycrystalline silicon ring resonator photodiodes in a bulk complementary metal-oxide-semiconductor process. *Opt. Lett.* **39**, 1061–1064 (2014).
- Atabaki, A. H., Meng, H., Alloatti, L., Mehta, K. K. & Ram, R. J. High-speed polysilicon CMOS photodetector for telecom and datacom. *Appl. Phys. Lett.* **109**, 111106 (2016).
- Nishi, Y. & Doering, R. *Handbook of Semiconductor Manufacturing Technology* (CRC Press, 2000).
- Soref, R. & Bennett, B. Electrooptical effects in silicon. *IEEE J. Quantum Electron.* **23**, 123–129 (1987).
- Ackert, J. J. et al. 10 Gbps silicon waveguide-integrated infrared avalanche photodiode. *Opt. Express* **21**, 19530–19537 (2013).
- Settaluri, K. T. et al. Demonstration of an optical chip-to-chip link in a 3D integrated electronic-photonic platform. In *41st European Solid-State Circuits Conf. (ESSCIRC)* 156–159 (IEEE, 2015).
- Sun, C. et al. A 45 nm CMOS-SOI monolithic photonics platform with bit-statistics-based resonant microring thermal tuning. *IEEE J. Solid-State Circ.* **51**, 893–907 (2016).
- Yang, S. et al. 10 nm high performance mobile SoC design and technology co-developed for performance, power, and area scaling. In *Symp. on VLSI Technology T70–T71* (IEEE, 2017).

Acknowledgements We thank E. Timurdogan and M. Byrd for help with high-speed photonic measurements, and A. Nikiforov of the Boston University Photonics Center for assistance with sample imaging and analysis. This work was supported by DARPA POEM award HR0011-11-C-0100 and contract HR00111190009, led by J. Shah. We also acknowledge support from the Microsystems Technology Laboratories (MTL) at MIT, the Berkeley Wireless Research Center, the Center for Nanoscale Sciences and Engineering and the

Packard Foundation. The views expressed are those of the authors and do not reflect the official policy or position of the DoD or the US Government.

Author contributions A.H.A., S.M. and F.P. all contributed equally to this work. A.H.A. designed and tested avalanche photodetectors, tested photonic devices during process development and designed WDM receiver rows. S.M. designed the mixed-signal electronics for transceiver chiplets, tested optical transceivers and electronics, and performed top-level assembly of electronics and photonics on the chip. F.P. designed and tested optical modulators, designed WDM transmitter rows, and performed top-level assembly of photonics in the transceiver chips. A.H.A., S.M. and F.P. all contributed to the chip verification. H.G. tested optical modulators and loss test structures on full flow wafers. J.N. designed and tested the original version of grating couplers. L.A. developed the CAD infrastructure for the photonics layout and chip verification, and contributed to the layout of waveguide test structures. M.T.W. designed and contributed to the layout of optical passive devices. C.S. contributed to the design of WDM optical transceivers and thermal tuning circuits. S.A.K. ran process experiments for propagation loss improvement, performed CMOS and photonics fabrication compatibility optimization, fabricated full flow wafers, and performed inline electrical and optical testing. H.M. contributed to the design of avalanche photodetectors. K.A.K. tested grating couplers on full-flow wafers.

I.W. performed device metrology and post processing of optical waveguide and resonator loss data. B.Z. performed grating coupler simulations and contributed to the new grating coupler designs. A.K. analysed device metrology and measurement data for studying and optimizing photonic performance. C.V.B. developed the 65-nm CMOS-compatible photonics design kit, designed methodology for the in-line electrical tests, metrology and optical photonics testing, worked with the mask house with building the photolithography masks and managed the overall process development and wafer fabrication. M.A.P., V.M.S. and R.J.R. supervised the project.

Competing interests C.S., M.T.W., R.J.R., M.A.P. and V.M.S. are involved in developing silicon photonic technologies at Ayar Labs.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0028-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.H.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Chip implementation. Photonic device layouts were developed and drawn in Cadence Virtuoso (an industry-standard design tool for frontend electronics in conjunction with mixed-signal electronics³⁴). Digital electronics were implemented using a combination of digital synthesis and place and route tools from Cadence. All photonic and electronic designs conform to the 65-nm CMOS technology manufacturing rules (more than 5,000 rules). New design rules were added to the original CMOS rules for the new photonics masks that were added to the process. The most critical mask rules with the introduction of photonics into the process are density rules. This is important because photonics and electronics occupy separate regions on the chip, but their respective masks have to maintain a certain maximum and minimum density of shapes across the whole design area. These density rules were met by custom fill shapes designed by our team. Density fill shapes can be seen in the scanning electron micrographs in Fig. 1c. The physical design verification was performed using Mentor Graphics Calibre (https://www.mentor.com/products/ic_nanometer_design/verification-signoff/).

Fabrication. Designs were fabricated on 300-mm wafers in the fabrication facility at Colleges for Nanoscale Sciences and Engineering, State University of New York, Albany, New York. The photonics passive-only wafers (partial flow) were fabricated on silicon wafers with 1.5- μm -thick SiO_2 under-cladding blankets with the whole CMOS backend dielectric stack as the over-cladding. Partial-flow wafers were fabricated for photonics process optimization before integration with electronics. For full-flow wafers (passive and active photonics with electronics), the deep photonic trench is first fabricated by etching the trench in the silicon substrate and filling it with SiO_2 by chemical vapour deposition followed by a planarization step. At this point, the wafer goes through the CMOS frontend process up to the source and drain formation. Photonic device fabrication is then followed by the deposition, annealing and planarization of a 220-nm photonic polysilicon layer. Using two reactive ion etching steps, one full (etching the entire 220 nm depth of polysilicon film) and one partial (120 nm deep), strip and ridge photonic structures are formed. This is followed by the photonic mid-level doping implants. From here, electronics and photonics share the rest of the fabrication process, including the high-doping implants (for the transistor source and drain, and the photonic modulator and detector ohmic contacts), nitride liners (silicide block, and etch-stop for the first via), silicide formation and metallization. There are a total of seven metal interconnect layers in this process, with the first four having a lithography resolution of less than 100 nm.

Each wafer quadrant on full-flow wafers received a separate mid-level doping implant concentration for photonic active components (modulators and detectors) of $[1, 2, 3, 6] \times 10^{18} \text{ cm}^{-3}$. Owing to the presence of a large density of defects in polysilicon, the carrier activation occurs only after the majority of defect states are occupied, whose onset occurs for a doping concentration³⁵ of roughly 10^{18} cm^{-3} . This necessitates careful optimization of photonic mid-level P and N doping concentrations to balance loss, modulator efficiency and device series resistance, which affects the speed of both modulators and detectors. By using a separate doping concentration in each quadrant of the wafer, we tested the performance of modulators and detectors as a function of doping concentration. From the results of doping splits in an earlier fabrication run for optimizing the photonics process, we expected the optimal doping concentration to be close to $3 \times 10^{18} \text{ cm}^{-3}$. However, in the full-flow run, owing to an increase in optical loss caused by polishing residues, microring Q -factors dropped by a factor of two. This required larger wavelength shifts in modulators to compensate for the broadened resonance lineshape to achieve the same level of modulation depth. Therefore, we observed the best overall performance for a P and N implant concentration of $6 \times 10^{18} \text{ cm}^{-3}$. The results presented in this paper are for devices receiving this implant concentration.

Fabrication results. The mask density rules ensure that material density during polishing, etching and lithography is within an acceptable range over the entire reticle and wafer to eliminate pattern-dependent results and achieve a high fabrication yield. Nevertheless, the maximum density range for each layer (polysilicon, metals, and so on) is desirable for more design flexibility. In this fabrication run, we faced unforeseen issues with photonic trench planarization, owing to a large density gradient of photonic trenches across the reticle field. This caused dielectric residues on the wafer after photonic trench planarization. We also experienced metal residues after the fabrication of the first via contact. Both of these issues led to a factor-of-two degradation in the passive photonic performance in full-flow runs (20 dB cm^{-1} versus 10 dB cm^{-1} for waveguide loss, and 10,000 versus 20,000 for microring Q -factor at 1,300 nm). Both of these issues were resolved through modified design rules, and optimized fabrication processes for the next fabrication run.

Optical testing. Tunable lasers from Agilent Technologies and Santec were used for the optical characterization. Standard single-mode fibres (SMF28) were used to couple light into and out of the chip using grating couplers. The width of the grating couplers is matched to the mode size of the SMF28 fibre. We used 3-axis positioner stages (Thorlabs NanoMax) to position and align fibres over the grating couplers of the test sites. Minimum fibre-to-coupler insertion loss was achieved by angling the

fibres at 15° off-normal from the surface of the chip. Waveguide losses were estimated by measuring transmitted optical power for four different waveguide lengths (50 μm , 1 mm, 5 mm and 15 mm), and fitting the propagation loss to the transmission measurements. Microring Q -factors were estimated by fitting a Lorentzian function to microrings at close to the critical coupling condition (the condition for zero transmission at resonance). Fibre-to-chip grating coupler efficiencies were extracted from the transmission measurement data, by fine-tuning the fibre angle at the input and output to achieve minimum transmission loss and subtracting the waveguide loss connecting the two identical grating couplers at the input and output of the test structure. Electro-optical frequency response (S_{21}) of optical modulators and photodetectors were measured using an Agilent Vector Network Analyzer (VNA, 8722D). For modulator bandwidth testing, the modulator was driven by VNA Channel 1 with the bias voltage applied using a Bias Tee (SHF BT 65), and the modulator output was detected by a high-bandwidth photodiode (Discovery DSC30-3-2010) which was connected to Channel 2 of the VNA for S_{21} measurement. For photodetector bandwidth testing, VNA Channel 1 was used to drive an external lithium niobate modulator (JDSU 10022054), and the photodetector was biased using SHF BT 65, whose radio-frequency output was connected to VNA Channel 2 for S_{21} measurement. We used high-bandwidth radio-frequency probes (Cascade Infinity, 50 μm pitch) for high-speed testing. Eye diagrams for the modulators and detectors were obtained using a similar setup, with a pseudorandom binary sequence pattern generator (Picosecond Programmable Pattern Generator, SDG Model 12072) and a high-bandwidth oscilloscope (Agilent Technologies 86108B Precision Waveform Analyzer). The microring photodetector responsivity was measured by dividing the device photocurrent by the input optical power, which was estimated by measuring the optical power in the input fibre before entering the chip and accounting for the fibre-to-chip grating coupler, and excess waveguide losses.

Electrical testing. Chips were assembled in ceramic packages (CPG20809) with 100 wirebond connections. These packages were plugged into a socket on a host printed circuit board, which delivers supplies, bias signals, the high-speed clock (from an Agilent 81142 A pulse generator), and scan control signals from an Opal-Kelly FPGA. Scan commands for each measurement are set in Python scripts on a computer and then sent to the FPGA to configure the chip for each particular experiment. To read out the ring-oscillator frequencies, the output of the oscillator is fed into an asynchronous digital divider (divide by 8) and the divided clock runs a digital counter block. The oscillator's frequency was then estimated by scanning out the counter's value. The transmitter's eye diagram was captured via an external Ortel photoreceiver with 10-GHz bandwidth on a digital communication analyser (DCA) oscilloscope by running the on-chip pseudorandom binary sequence modules at a 5-GHz external clock frequency. On-chip clock adjustment circuits composed of a duty-cycle corrector and a delay line were used to synchronize the timing of different transceivers. The receiver's bit-error-rate test was performed by first programming a KC705 Xilinx FPGA with the same pseudorandom binary sequence coefficients used for our on-chip bit-error-rate checkers. The output of the FPGA was then amplified using a high-voltage modulator driver (JDSU H301), which then drives an external JDSU MZI optical modulator. Modulated light is amplified using a semiconductor optical amplifier (Thorlab BOA1130) and is coupled into the chip. The bit-error-rate bathtub curves in Fig. 4f show the measured bit-error rate at each time delay point between the clock fed into the chip and the FPGA reference clock.

Device design and discussion. The reticle area was divided into two main sections: the test device section with photonic and electrical test structures and the electronic-photonic microsystem section (WDM chiplets). The test area included waveguide loss, microring Q -factor, grating coupler efficiency, and sheet resistance test structures as well as individual modulators and detectors. A variety of parameters in every device was swept to find optimal designs and to extract information about the quality of the fabrication (for example, estimating surface and sidewall roughness, doping activation, and so on). Photonic test devices were designed for both 1,300 nm and 1,550 nm wavelengths. Overall, approximately 1,000 test devices were laid out on the reticle. The microsystems included 6 transceiver chiplets, each 4.8 mm \times 5 mm. Since the same set of masks was used for process optimization and system implementation, there were some uncertainties about the performance of photonic passives and actives. This required sweeping modulator and detector parameters in the WDM transceiver designs to cover a large enough range of device performance.

All photonic components were designed using two etch steps (partial and full). The thickness of the polysilicon layer (220 nm) and the depth of the partial etch (120 nm) were chosen to optimize the overall performance of the whole platform, including the efficiency of grating couplers, radiation loss of the microring resonators, and series resistance of the modulators and detectors. The optimal width of the single-mode waveguides and diameter of the high- Q microrings were around 450 nm and 15 μm , respectively, for 1,300 nm operation. The high doping regions for ohmic contacts were about 1 μm away from the centre of the ridge waveguides

to avoid free-carrier loss. The width of the intrinsic region for the photodiode is 800 nm in the reported device. Microheaters were incorporated in the microring modulators and detectors for tuning their resonances to the desired laser wavelength. We used silicided P+ polysilicon resistors for the microheaters. The fabricated microheaters had a resistance of about 100 Ω .

The thickness of the photonic trench was optimized to maximize the grating coupler efficiency. A thickness of approximately 1.5 μm results in constructive interference of the main beam diffracted upwards with the beam diffracted downward and reflected from the surface of the silicon substrate. This improved the directionality and coupling efficiency of the grating couplers.

Device micrographs. The micrographs of the chip and discrete photonic components in Fig. 2 are taken from the back side of the die, subsequent to complete removal of the silicon substrate using XeF_2 gas after mounting the die on a carrier substrate.

Improvement of photonic performance. We have already taken the necessary steps to improve the waveguide loss in our full-flow wafers by fixing the photonic isolation planarization issue on the next fabrication run. Hence, we expect to achieve the same passive photonics performance in our fully integrated wafers as that reported in partial-flow runs in this work. This means improvements of a factor of two in waveguide loss (10 dB cm^{-1}) and microring Q-factor (20,000), simply by resolving the photonic isolation planarization issue.

The loss improvement will also affect the performance of active photonics, resulting in a higher quantum efficiency for photodetectors by reducing the fraction of photons lost by scattering. We expect a 50%–100% improvement in detector responsivity (0.15–0.2 A W^{-1} in the linear mode, and 2–2.6 A W^{-1} with avalanche gain). The improvement in the microring modulator Q-factor leads to a sharper resonance feature, which consequently reduces the drive voltage and lowers the transmitter power consumption.

The waveguide loss can be improved even further to about 6 dB cm^{-1} across the entire telecom and datacom bands (1,250–1,700 nm) by optimizing the polysilicon film. At present, waveguide loss at shorter wavelengths (1,300 nm) is twice as high as at 1,550 nm (Fig. 1). The similarity of the detector quantum efficiency (11%) at 1,300 nm and 1,550 nm suggests that the scattering and absorption losses are increasing proportionately at shorter wavelengths. As optical modes at shorter wavelengths are more confined in the waveguide core and exhibit less scattering by the sidewall roughness, the dominant mechanism for the increase in scattering loss at shorter wavelengths is most probably the increase in scattering by polysilicon grain boundaries. This source of scattering can be reduced by optimizing polysilicon deposition and anneal conditions such that the scattering correlation length is reduced.

The bandwidth of the modulators and detectors can also be improved by optimizing the doping profiles. In the present run, we employed the same source and drain implants for the low-resistance regions (P++ and N++) of our modulators and detectors. These implants have concentrations that are designed for the shallow source and drain of deeply scaled transistors and are not high enough to provide low series resistance for the 100-nm-thick polysilicon regions in our photonic devices. By using dedicated implant masks and high implant dosages, and by further optimizing dimensions in all doping profiles, we expect to improve the bandwidth of modulators by 25%–50% to 20–24 GHz based on the estimates for

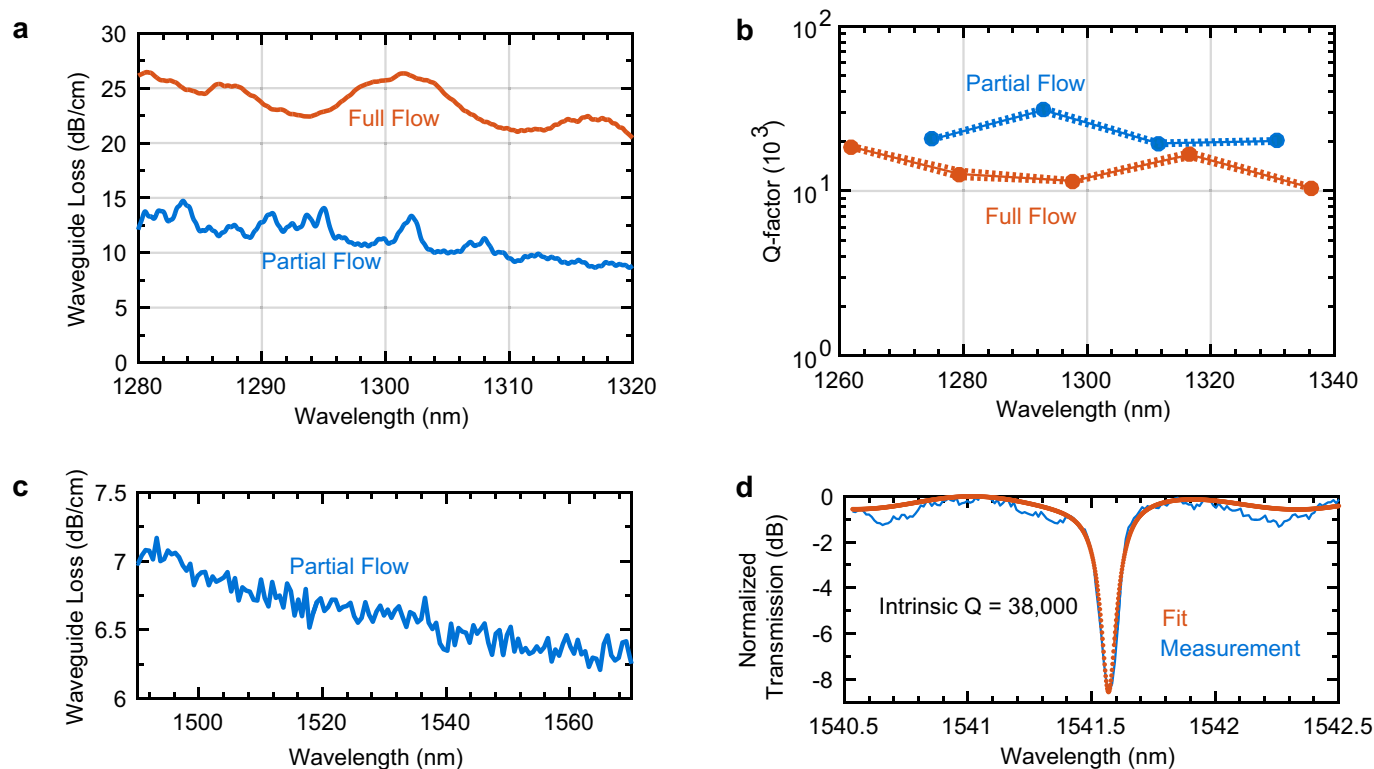
the contributions of mid-level (P and N) and high-level doping (P++ and N++) regions to the series resistance. For the detectors, we also expect an improvement in bandwidth by reducing the width of the intrinsic region from 800 nm to reduce the transit time, which is at present limiting the speed of the device. We expect that a 50% reduction of the intrinsic region width to 400 nm would not substantially affect the Q-factor and responsivity of detectors, while reducing the depletion width and transit time. A careful optimization of the intrinsic region width combined with the reduction of the RC time constant (resistance \times capacitance) by adjusting the implant conditions can improve the bandwidth of detectors by 25%–50% to 14–16.5 GHz in the linear mode and to 10–12 GHz in the avalanche mode.

We also expect an improvement in the grating coupler efficiency in the next fabrication run. In the present run, a 30-nm photolithography bias caused the dimensions of the grating couplers to be smaller than the nominal design values. This caused the grating coupler efficiency to drop from -1.8 dB in design to -5 dB in the fabricated devices. We are addressing this issue by optimizing the lithography step and pre-biasing the photolithography mask. Also, by using a nonuniform grating design³⁶, we expect to improve the mode matching of the grating coupler to the Gaussian mode profile of the optical fibre. This will enable us to improve the coupler efficiency to below -1 dB.

Electronic–photonic systems on glass. The present work was aimed at integrating photonics into bulk CMOS technologies. However, a fully functional deposited photonic platform on glass transcends any one particular substrate or application. All of our partial-flow photonics-only silicon wafers were covered by a blanket of 1.5 μm -thick plasma-enhanced chemical vapour deposition silicon oxide, on which we have fabricated optical waveguides, resonators, modulators and photodetectors. These thin-film integrated photonic devices, along with the thin-film transistors that are currently used in display panels, could enable electronic–photonic systems on glass. These systems can be fabricated on low-cost large-area substrates such as metal foils, transparent glass or even flexible substrates as long as they are covered with roughly 1 μm of glass. Such a platform can enable a variety of new systems and applications that current electronic–photonic technologies cannot address owing to substrate size or cost limitations. For example, several space and astrophysics applications, such as laser communications and astronomical spectroscopy, require large-area optics and detectors. Also, many optical phased array applications (lidars, augmented reality headsets, and so on) could benefit greatly from large-area integrated photonic circuits. An electronic–photonic platform on glass, enabled by the deposited polysilicon photonics demonstrated in this work, could address these application areas. The performance of photonics on this platform would be similar to the devices we have reported on partial-flow wafers in this paper.

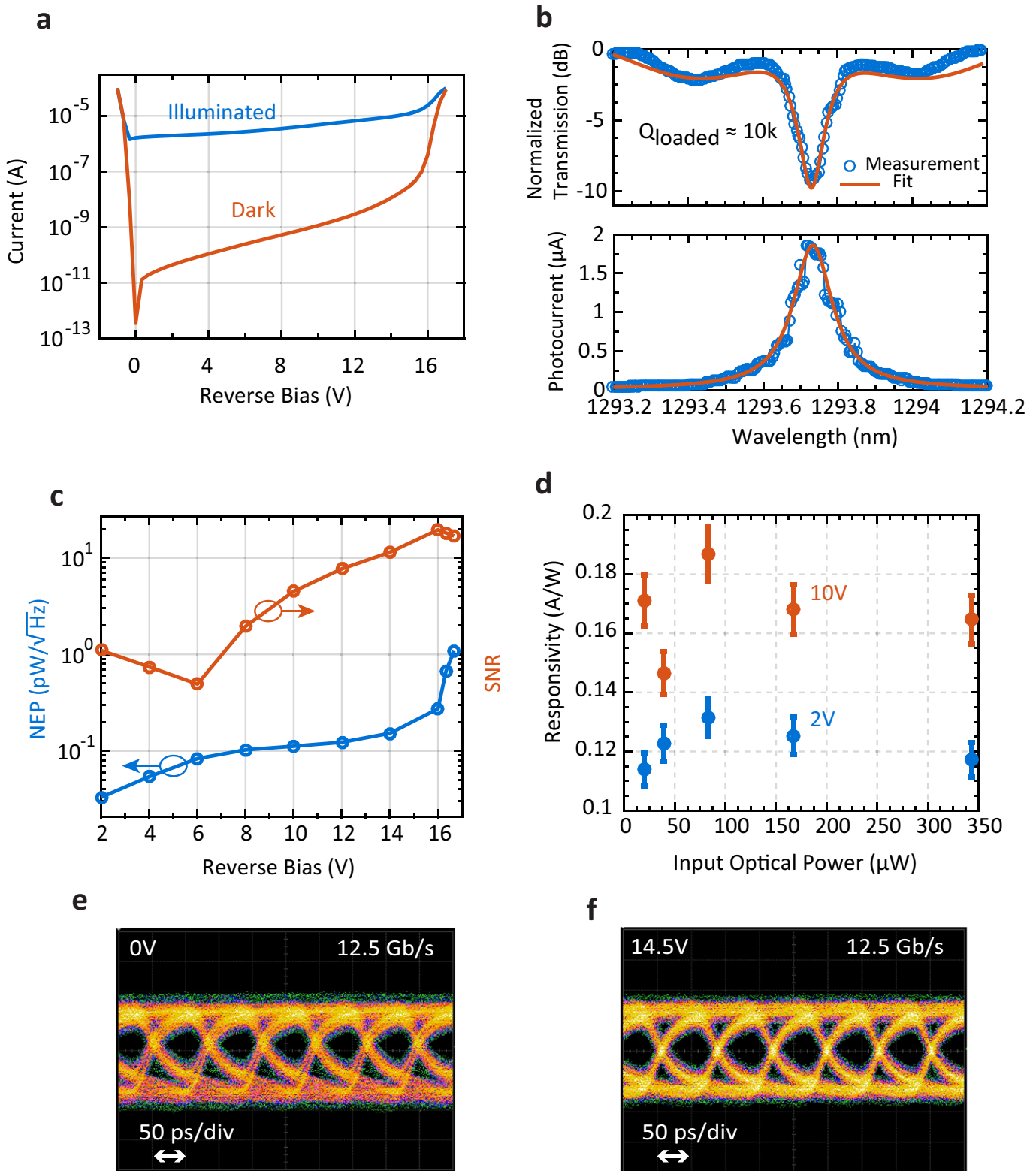
Data availability. The main data supporting the findings of this study are available within the article. Extra data are available from the corresponding author upon request.

34. Alloatti, L., Wade, M., Stojanovic, V., Popovic, M. & Ram, R. Photonics design tool for advanced CMOS nodes. *IET Optoelectron.* **9**, 163–167 (2015).
35. Seto, J. Y. W. The electrical properties of polycrystalline silicon films. *J. Appl. Phys.* **46**, 5247 (1975).
36. Notaros, J. et al. Ultra-efficient CMOS fiber-to-chip grating couplers. In *Optical Fiber Communications Conf. Exhib. (OFC)* 1–3 (IEEE, 2016).



Extended Data Fig. 1 | Passive photonic performance at 1,300 nm and 1,550 nm. a, Waveguide propagation loss at 1,300 nm. Waveguide loss drops with wavelength because of a combination of lower absorption

and scattering by polysilicon. **b**, Q-factor of a 15- μm -diameter microring resonator. **c**, Waveguide propagation loss at 1,550 nm. **d**, One resonance of a 17- μm -diameter microring near 1,540 nm with a Q-factor of 38,000.



Extended Data Fig. 2 | Polysilicon avalanche photodiode. **a**, Current-voltage curve of the microring photodiode under dark and illumination for an input optical power of $20\mu\text{W}$. Dynamic range is about 60 dB and about 10 dB at 0 V and 16 V, respectively. **b**, One microring photodetector resonance (top) and the corresponding photo-current (bottom) as the wavelength is swept across the resonance. The loaded Q -factor (Q_{loaded}) of the microring is about 10,000. The fit is obtained through least-squares optimization of a model that includes a Lorentzian resonance for the microring and accounts for the reflections from the end facets of the chip to model the Fabry–Pérot resonances observed in the transmission curve. **c**, Noise equivalent power (NEP, blue curve) of the photodiode

estimated from the dark-current shot noise, which dominates the detector noise. Avalanche gain is 13 at 16 V bias, with a noise equivalent power of $0.27\text{ pW Hz}^{-1/2}$. The simulated signal-to-noise ratio (SNR) (red curve) at the output of the optical receiver, assuming an optical signal of $1\mu\text{W}$, and a receiver circuit input-referred noise spectral density of $1\text{ pA Hz}^{-1/2}$. **d**, The responsivity of the photodetector versus input optical power, showing minimal power dependency. The error bar is estimated based on a $\pm 5\%$ error in estimating the optical power in the waveguide before coupling into the detector. This error comes from variations in fibre to chip coupling efficiency owing to fibre-grating coupler misalignment. **e**, **f**, Eye diagrams at 12.5 Gb s^{-1} at 0 V and 14.5 V reverse bias.

A library of atomically thin metal chalcogenides

Jiadong Zhou^{1,15}, Junhao Lin^{2,15*}, Xiangwei Huang³, Yao Zhou⁴, Yu Chen⁵, Juan Xia⁵, Hong Wang¹, Yu Xie⁶, Huimei Yu⁷, Jincheng Lei⁶, Di Wu^{8,9}, Fucui Liu¹, Qundong Fu¹, Qingsheng Zeng¹, Chuang-Han Hsu^{8,9}, Changli Yang^{3,10}, Li Lu^{3,10}, Ting Yu⁵, Zexiang Shen⁵, Hsin Lin^{8,9,11}, Boris I. Yakobson⁶, Qian Liu⁴, Kazu Suenaga², Guangtong Liu^{3*} & Zheng Liu^{1,12,13,14*}

Investigations of two-dimensional transition-metal chalcogenides (TMCs) have recently revealed interesting physical phenomena, including the quantum spin Hall effect^{1,2}, valley polarization^{3,4} and two-dimensional superconductivity⁵, suggesting potential applications for functional devices^{6–10}. However, of the numerous compounds available, only a handful, such as Mo- and W-based TMCs, have been synthesized, typically via sulfurization^{11–15}, selenization^{16,17} and tellurization¹⁸ of metals and metal compounds. Many TMCs are difficult to produce because of the high melting points of their metal and metal oxide precursors. Molten-salt-assisted methods have been used to produce ceramic powders at relatively low temperature¹⁹ and this approach²⁰ was recently employed to facilitate the growth of monolayer WS₂ and WSe₂. Here we demonstrate that molten-salt-assisted chemical vapour deposition can be broadly applied for the synthesis of a wide variety of two-dimensional

(atomically thin) TMCs. We synthesized 47 compounds, including 32 binary compounds (based on the transition metals Ti, Zr, Hf, V, Nb, Ta, Mo, W, Re, Pt, Pd and Fe), 13 alloys (including 11 ternary, one quaternary and one quinary), and two heterostructured compounds. We elaborate how the salt decreases the melting point of the reactants and facilitates the formation of intermediate products, increasing the overall reaction rate. Most of the synthesized materials in our library are useful, as supported by evidence of superconductivity in our monolayer NbSe₂ and MoTe₂ samples^{21,22} and of high mobilities in MoS₂ and ReS₂. Although the quality of some of the materials still requires development, our work opens up opportunities for studying the properties and potential application of a wide variety of two-dimensional TMCs.

Figure 1 proposes a general picture for the synthesis of two-dimensional (2D) TMCs using the chemical vapour deposition method,

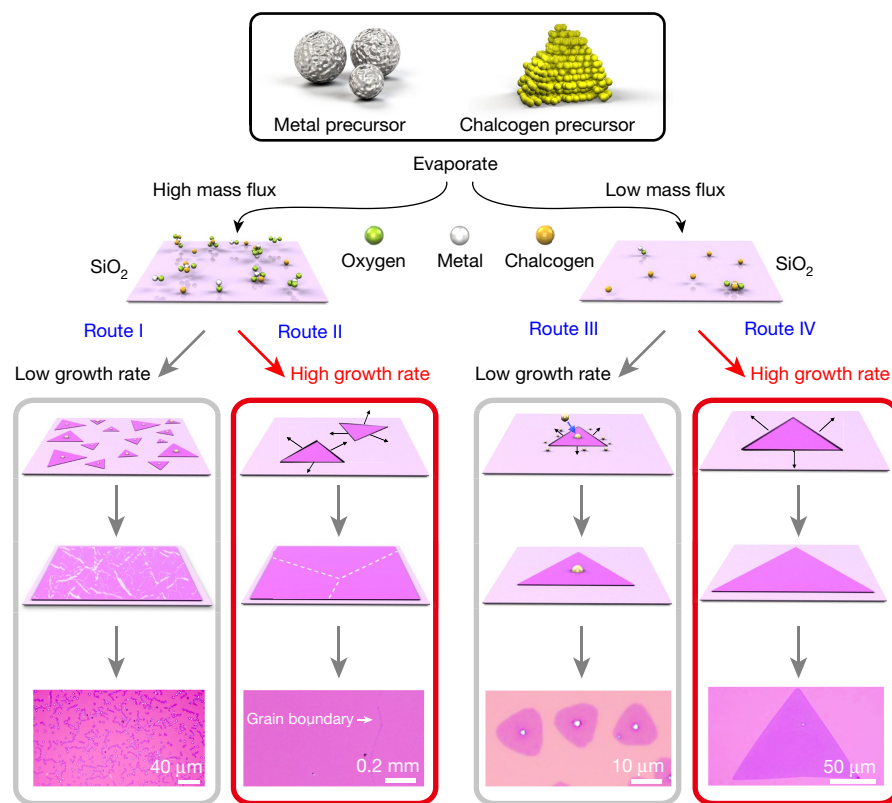


Fig. 1 | Flow chart of the general growth process for the production of TMCs by the chemical vapour deposition method. The growth of 2D TMCs can be classified into four routes based on different mass flux of metal precursor and growth rate. High mass flux of metal precursor offers the opportunity to synthesize large-scale continuous monolayer polycrystalline films with small (route I) or large (route II) domains depending on the growth rate. On the other hand, low mass flux of metal precursor results in discrete single-crystalline monolayers with different sizes. Low growth rate leads to small crystal size with atom clusters decorated in the centre and edge of the monocrystal (route III), while high growth rate gives rise to large monocrystals (route IV).

¹Center for Programmable Materials, School of Materials Science and Engineering, Nanyang Technological University, Singapore, Singapore. ²National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan. ³Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China. ⁴The State Key Laboratory of High Performance Ceramics and Superfine Microstructure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai, China. ⁵Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore. ⁶Department of Materials Science and NanoEngineering and Department of Chemistry, Rice University, Houston, TX, USA. ⁷School of Materials Science and Engineering, East China University of Science and Technology, Shanghai, China. ⁸Centre for Advanced 2D Materials and Graphene Research Centre, National University of Singapore, Singapore, Singapore. ⁹Department of Physics, National University of Singapore, Singapore, Singapore. ¹⁰Collaborative Innovation Center of Quantum Matter, Beijing, China. ¹¹Institute of Physics, Academia Sinica, Taipei, Taiwan. ¹²Centre for Micro-/Nano-electronics (NOVITAS), School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore. ¹³CINTRA CNRS/NTU/THALES, UMI 3288, Research Techno Plaza, Singapore, Singapore. ¹⁴Environmental Chemistry and Materials Centre, Nanyang Environment and Water Research Institute, Singapore, Singapore. ¹⁵These authors contributed equally: Jiadong Zhou, Junhao Lin. *e-mail: lin.junhao.stem@gmail.com; gtlui@phy.ac.cn; z.liu@ntu.edu.sg

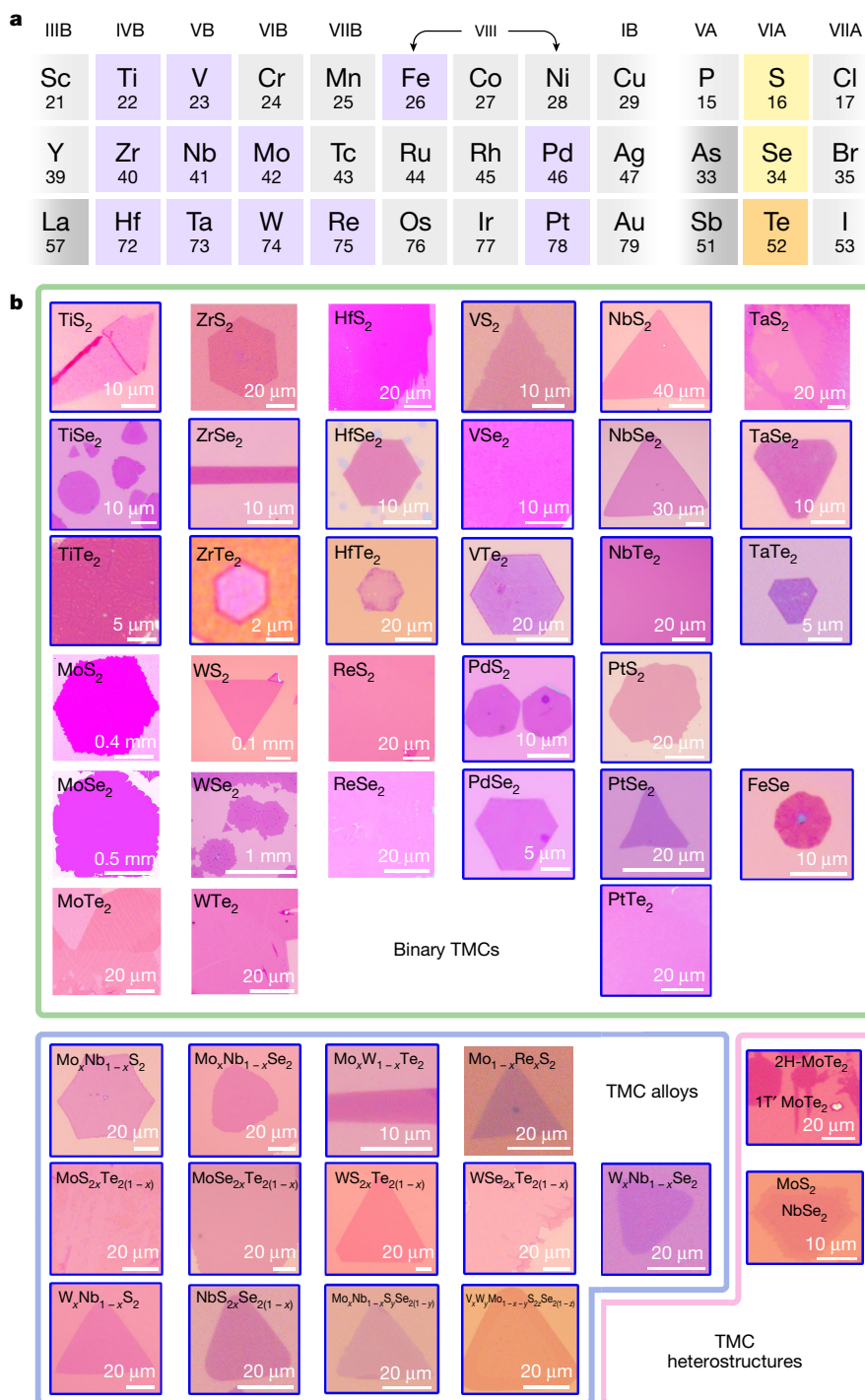


Fig. 2 | The transition metals and chalcogens used, and optical images of the resulting 47 different atomically thin TMCs and heterostructures. a, Overview of metals (highlighted in purple) and chalcogens (highlighted in yellow and orange) that can form layered sulfides, selenides and tellurides. **b,** Optical images of 47 TMCs synthesized using our method: binary 2D crystals containing Mo (MoS₂, MoSe₂, MoTe₂), W (WS₂, WSe₂, WTe₂), Re (ReS₂, ReSe₂), Ti (TiS₂, TiSe₂, TiTe₂), Zr (ZrS₂, ZrSe₂, ZrTe₂), Hf (HfS₂, HfSe₂, HfTe₂), V (VS₂, VSe₂, VTe₂), Nb (NbS₂, NbSe₂, NbTe₂), Ta (TaS₂, TaSe₂, TaTe₂), Pt (PtS₂, PtSe₂, PtTe₂),

Pd (PdS₂, PdSe₂) or Fe (FeSe); the ternary alloys MoS_xTe_{2-x}, MoSe_xTe_{2-x}, WS_xTe_{2-x}, WSe_xTe_{2-x}, NbS_xSe_{2-x}, Mo_xNb_{1-x}S₂, Mo_xNb_{1-x}Se₂, Mo_{1-x}Re_xS₂, W_xNb_{1-x}S₂, W_xNb_{1-x}Se₂ and Mo_xW_{1-x}Te₂; the quaternary alloy Mo_xNb_{1-x}S_{2-y}Se_{2(1-y)}; the quinary alloy V_xW_yMo_{1-x-y}S_{2z}Se_{2(1-z)}; and the 1 T' MoTe₂-2 H MoTe₂ in-plane and MoS₂-NbSe₂ vertically stacked heterostructures. TMCs that have not been previously synthesized are outlined in blue. Detailed characterizations of the as-grown 2D materials are shown in Supplementary Information.

based on the competition between the mass flux of the metal precursors and the reaction rate of the domains. The mass flux determines the amount of metal precursors involved in the formation of the nucleus and the growth of domains, whereas the growth rate dominates the grain size of the as-grown films. At high mass flux, low growth rate results in a monolayer polycrystalline film (route I) with small grains,

and high growth rate tends to form continuous monolayer films with large grains of up to millimetres in size (route II)²³. On the other hand, at low mass flux, low growth rate promotes the formation of small flakes. Tiny nuclei are often observed at the centre of the flakes²⁴, suggesting that the extra adatoms or atom clusters will consistently attach to an existing nucleus or to the edge during growth (route III). A high

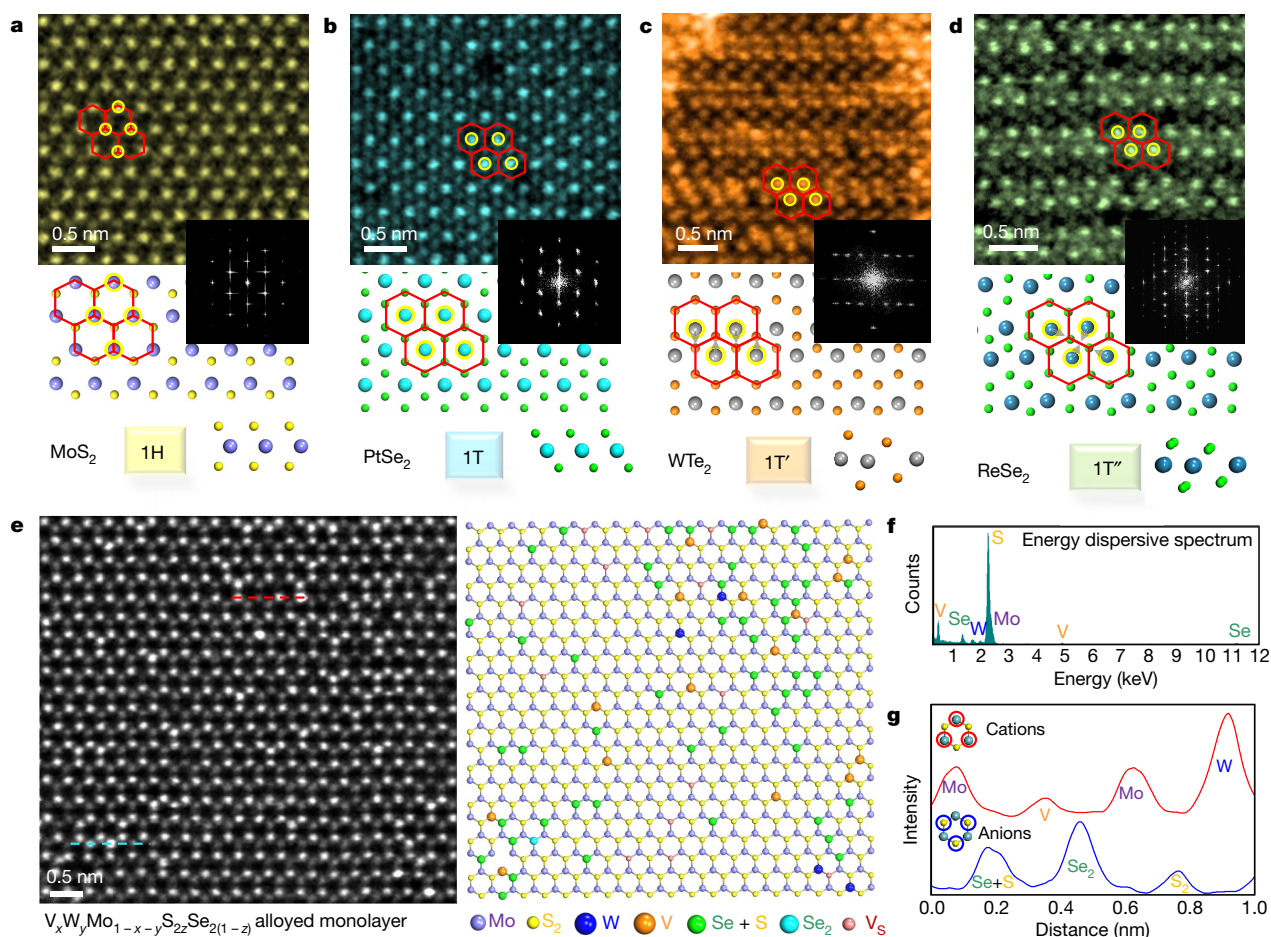


Fig. 3 | Atomic-resolution STEM images of representative monolayer materials in different phases. a–d, MoS₂ in the 1H phase (a), PtSe₂ in the 1T phase (b), WTe₂ in the 1T' phase (c) and ReSe₂ in the 1T'' phase (d), with their corresponding fast Fourier transform patterns and atomic structural models. e, STEM image (left) of the quinary monolayer alloy

$V_xW_yMo_{1-x-y}S_{2z}Se_{2(1-z)}$, with the corresponding atomic model from atom-by-atom intensity mapping (right). f, EDS spectra of the alloyed monolayer, confirming its chemical composition. g, Line intensity profiles along the highlighted red (cation) and blue (anion) dashed lines in e, indicating the different intensity of each chemical species.

reaction rate preferentially produces a monolayer of individual large 2D single crystals (route IV)²⁵.

Unfortunately, many TMCs, such as those based on Nb, Pt and Ti, are very difficult to produce because their metal or metal oxide precursors have high melting points and low vapour pressure, which leads to very low mass flux and limits the reaction. Molten salts can increase mass flux by reducing the melting points of the metal precursors and forming oxychlorides via reaction with some metal oxides, thus increasing the rate of the reaction. Using this method, we have synthesized a library of 2D TMCs (shown in Fig. 2).

Figure 2a shows a schematic of the periodic table, highlighting the chemical combination of all the 2D TMC atomic layers produced, formed between 12 transition metals (purple) and three chalcogens (yellow). Synthetic recipes and reaction conditions are illustrated in Methods and are summarized in Supplementary Fig. 1 and Supplementary Table 1. Figure 2b shows optical images of 47 2D TMC compounds with morphologies of triangles, hexagons, ribbons and films, including 32 binary crystals from group IVB (Ti, Zr and Hf) to group VIII (Pd and Pt), 13 alloyed TMCs (containing ternary, quaternary and even quinary compounds), which are important for universal bandgap engineering and heterogeneous catalysis^{26–29}, and two heterostructure TMCs (vertically stacked MoS₂–NbSe₂ and in-plane 1T' MoTe₂–2H MoTe₂). The 2D TMCs that had not previously been synthesized are outlined in blue. Detailed characterizations of all 2D TMCs and heterostructures are presented in Supplementary Figs. 2–11.

The atomic structures and chemical compositions of the as-synthesized 2D crystals and compounds are further revealed by atom-resolved scanning transmission electron microscopy (STEM)

imaging, energy-dispersive X-ray spectroscopy (EDS) and electron energy loss spectroscopy (EELS). The atomic structures of most 2D crystals can be classified into four types: (1) the trigonal prismatic 1H phase; (2) the undistorted 1T phase with the metal atom located at the centre of an octahedral unit; (3) the one-dimensional distorted 1T phase (called the 1T' phase), in which pairs of metal atoms move closer to each other perpendicularly, resulting in a quasi-one-dimensional chain-like structure consisting of distorted octahedral units; and (4) the two-dimensional distorted 1T phase (called the 1T'' phase), in which four nearby metal atoms move closer to each other to form a new unit cell, producing repeatable diamond-like patterns. The structural phase of each synthesized material can be determined by the Z-contrast STEM image. Figure 3a–d shows the representative materials for each phase with the corresponding atomic structural models; these materials are monolayer MoS₂, PtSe₂, WTe₂ and ReSe₂ for the 1H, 1T, 1T' and 1T'' phases, respectively. The patterns obtained by fast Fourier transform further indicate that the 1H and 1T phases maintain a hexagonal unit cell, whereas the 1T' phase forms a rectangular unit cell owing to one-dimensional metal-pair distortion and the 1T'' phase changes to a much larger hexagonal cell owing to the aggregation of four metal atoms into a new unit cell. A summary of different phases for each as-synthesized 2D material that has been examined is shown in Supplementary Fig. 12. Details of the atomic structure, EDS and EELS characterizations for each 2D crystal are given in Supplementary Figs. 13–27.

A STEM image of a quinary $V_xW_yMo_{1-x-y}S_{2z}Se_{2(1-z)}$ monolayer alloy is shown in Fig. 3e, where the chemical composition is verified by the EDS spectrum (Fig. 3f). The corresponding optical image

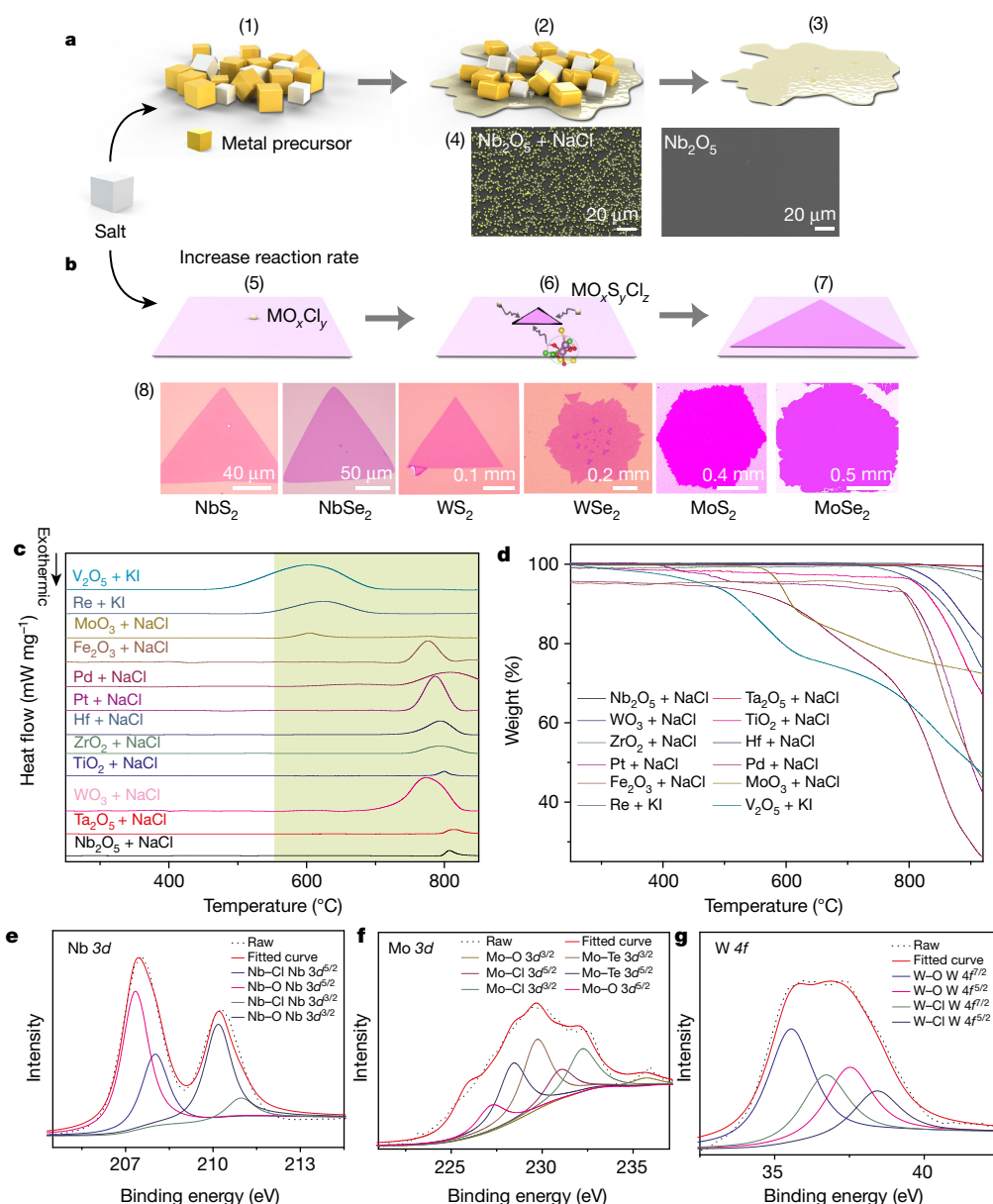


Fig. 4 | Reaction mechanism. **a, b**, Schematics of the reactions. Metal oxychlorides are formed, and these promote the reactions. Chalcogens are not shown here. (1)–(3) The proposed process of the added salt decreasing the melting point of the precursors. (4) SEM images of the Nb nucleus with (left) and without (right) added salt. **b**, (5)–(7) The growth process of the 2D atomic layer, with intermediate products. (8) Different single-crystalline monolayers with large size of monolayer and a growth time of less than 3 min. **c, d**, The TG-DSC curve of salts mixed with the metal sources. The melting points of the systems after adding salt are all

within the highlighted windows from 600 °C to 850 °C. **e–g**, XPS spectra revealing the existence of Cl bonds to other elements, resulting from the intermediate products during the synthesis of Nb-, Mo- and W-based 2D crystals. XPS spectra of Nb 3d, Mo 3d and W 4f are shown in **e, f** and **g**, respectively. Nb 3d^{3/2} and 3d^{5/2}, Mo 3d^{3/2} and 3d^{5/2} and W 4f^{5/2} and 4f^{7/2} are the core level energy states of Nb, Mo and W, respectively, all of which are well fitted with Gaussian peaks at energies indicating the bonding to Cl.

and Raman spectrum are shown in Supplementary Fig. 9. Different chemical species give rise to the distinct atomic contrast in the image. Combined with the intensity histogram analysis of the cation and anion sites (Supplementary Fig. 28), each atomic column can be directly associated with their chemical identities using the image contrast, as shown by the representative line intensity profile in Fig. 3g. The atom-by-atom mapping further confirms the successful synthesis of a quinary alloyed monolayer. We also observe superconductivity in monolayer NbSe₂ and MoTe₂ (Supplementary Figs. 29–31)^{21,22}, which is the realization of superconductivity in non-ultrahigh-vacuum-grown monolayer materials. Combined with the high mobility of monolayer MoS₂ and ReS₂ (Supplementary Figs. 32 and 33), these results indicate the high quality of the as-prepared 2D TMCs. We note that most of the as-synthesized materials show well controlled thickness and useful

attributes, but some, such as ZrTe₂, TiTe₂, HfTe₂ and the Pd-based ones, need to be further improved.

The growing mechanism of the salt-assisted chemical vapour deposition method is discussed in Fig. 4 and detailed in Supplementary Figs. 34–53). Figure 4a illustrates that salt can reduce the melting points of metal precursors, and thus make the reaction possible (see also Supplementary Fig. 35) to grow 2D TMCs. As an example, a comparison of the observed Nb nucleus with and without salt added is shown in Fig. 4a, indicating a high mass flux of the metal precursors promoted by the salt. In addition to decreases of the melting point, Fig. 4b shows that some metal oxides can react with salt to form metal oxychlorides, which evaporate at an appropriate temperature and facilitate the growth of the 2D TMCs. We obtained the melting points of the precursors for all binary 2D systems using thermogravimetry and

differential scanning calorimetry (TG-DSC) measurements. They all fell within the temperature window from 600 °C to 850 °C, as shown in Fig. 4c, which matches the temperature range in which the resulting materials grow. This is further supported by the thermogravimetry versus time curves in Fig. 4d. During the growth, coarsening forms a stable nucleus (Supplementary Fig. 38), then adatoms and atom clusters of chalcogen and metal attach to the edges of as-grown 2D monolayers and grow quickly owing to their high mobility (Supplementary Figs. 47 and 48). This growth process is supported by experimental evidence (Supplementary Figs. 49–51), and also previous reports³⁰ on MoS₂. This helps to produce millimetre-sized single-crystal 2D TMCs, such as the W-, Nb- and Mo-based TMCs (Fig. 4b). Notably, the growth time in our experiment is as short as 3 min and the growth rate is up to 8 $\mu\text{m s}^{-1}$ (Supplementary Figs. 39 and 40), owing to the high chemical activities of oxychloride during the reaction. To confirm the existence of metal oxychloride, the intermediate products are collected and analysed by X-ray photoelectron spectroscopy (XPS) during the synthesis of monolayer NbX₂, MoX₂ and WX₂ (X = S, Se, Te). The signals from M–Cl and M–O (M = W, Nb, Mo) bonds in Nb 3d, Mo 3d and W 4f (Fig. 4e–g) confirm the existence of the oxychloride compounds^{31–33} NbO_xCl_y, MoO_xCl_y and WO_xCl_y (see also Supplementary Figs. 52 and 53). Moreover, the formation of metal oxychloride is also corroborated by ab initio molecular dynamics simulations (Supplementary Figs. 36 and 37). The density functional theory calculations show that it is energetically more favourable to sulfurize metal oxychlorides than metal oxides.

In conclusion, we have demonstrated a universal salt-assisted chemical vapour deposition method for producing a 2D TMC library consisting of 47 compounds and heterostructures. Our work provides a swift way to produce good-quality 2D TMCs. Our understanding of the growth mechanism is greatly improved, suggesting ways to explore the extraordinary physical characteristics of these materials and their nanodevice applications.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0008-3>.

Received: 31 December 2016; Accepted: 26 January 2018;

Published online 18 April 2018.

- Zhang, Y. B., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Qian, X. F., Liu, J. W., Fu, L. & Li, J. Quantum spin Hall effect in two-dimensional transition metal dichalcogenides. *Science* **346**, 1344–1347 (2014).
- Xiao, D., Liu, G. B., Feng, W. X., Xu, X. D. & Yao, W. Coupled spin and valley physics in monolayers of MoS₂ and other group-VI dichalcogenides. *Phys. Rev. Lett.* **108**, 196802 (2012).
- Zeng, H. L., Dai, J. F., Yao, W., Xiao, D. & Cui, X. D. Valley polarization in MoS₂ monolayers by optical pumping. *Nat. Nanotechnol.* **7**, 490–493 (2012).
- Saito, Y., Nojima, T. & Iwasa, Y. Highly crystalline 2D superconductors. *Nat. Rev. Mater.* **2**, 16094 (2017).
- Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Radisavljevic, B. & Kis, A. Mobility engineering and a metal-insulator transition in monolayer MoS₂. *Nat. Mater.* **12**, 815–820 (2013).
- Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699–712 (2012).
- Roy, K. et al. Graphene-MoS₂ hybrid structures for multifunctional photoresponsive memory devices. *Nat. Nanotechnol.* **8**, 826–830 (2013).
- Lopez-Sanchez, O., Lembke, D., Kayci, M., Radenovic, A. & Kis, A. Ultrasensitive photodetectors based on monolayer MoS₂. *Nat. Nanotechnol.* **8**, 497–501 (2013).
- Zhan, Y. J., Liu, Z., Najmaei, S., Ajayan, P. M. & Lou, J. Large-area vapor-phase growth and characterization of MoS₂ atomic layers on a SiO₂ substrate. *Small* **8**, 966–971 (2012).
- van der Zande, A. M. et al. Grains and grain boundaries in highly crystalline monolayer molybdenum disulfide. *Nat. Mater.* **12**, 554–561 (2013).
- Lee, Y. H. et al. Synthesis of large-area MoS₂ atomic layers with chemical vapor deposition. *Adv. Mater.* **24**, 2320–2325 (2012).
- Lin, Y. C. et al. Wafer-scale MoS₂ thin layers prepared by MoO₃ sulfurization. *Nanoscale* **4**, 6637–6641 (2012).
- Elias, A. L. et al. Controlled synthesis and transfer of large-area WS₂ sheets: from single layer to few layers. *ACS Nano* **7**, 5235–5242 (2013).
- Lu, X. et al. Large-area synthesis of monolayer and few-layer MoSe₂ films on SiO₂ substrates. *Nano Lett.* **14**, 2419–2425 (2014).
- Huang, J. K. et al. Large-area synthesis of highly crystalline WSe₂ monolayers and device applications. *ACS Nano* **8**, 923–930 (2014).
- Park, J. C. et al. Phase-engineered synthesis of centimeter-scale 1T'- and 2H-molybdenum ditelluride thin films. *ACS Nano* **9**, 6548–6554 (2015).
- Kimura, T. in *Advances in Ceramics—Synthesis and Characterization Processing and Specific Applications* (ed. Sikalis, C.) Ch. 2 (InTech, London, 2011).
- Li, S. S. et al. Halide-assisted atmospheric pressure growth of large WSe₂ and WS₂ monolayer crystals. *Appl. Mater. Today* **1**, 60–66 (2015).
- Xi, X. X. et al. Ising pairing in superconducting NbSe₂ atomic layers. *Nat. Phys.* **12**, 139–143 (2016).
- Xi, X. X. et al. Strongly enhanced charge-density-wave order in monolayer NbSe₂. *Nat. Nanotechnol.* **10**, 765–769 (2015).
- Dumcenco, D. et al. Large-area epitaxial monolayer MoS₂. *ACS Nano* **9**, 4611–4620 (2015).
- Li, B. et al. Solid-vapor reaction growth of transition-metal dichalcogenide monolayers. *Angew. Chem. Int. Ed.* **55**, 10656–10661 (2016).
- Gong, Y. J. et al. Synthesis of millimeter-scale transition metal dichalcogenides single crystals. *Adv. Funct. Mater.* **26**, 2009–2015 (2016).
- Chen, Y. F. et al. Tunable band gap photoluminescence from atomically thin transition-metal dichalcogenide alloys. *ACS Nano* **7**, 4610–4616 (2013).
- Gong, Y. J. et al. Band gap engineering and layer-by-layer mapping of selenium-doped molybdenum disulfide. *Nano Lett.* **14**, 442–449 (2014).
- Lin, Z. et al. Facile synthesis of MoS₂ and Mo₂W_{1-x} triangular monolayers. *APL Mater.* **2**, 092514 (2014).
- Azizi, A. et al. Spontaneous formation of atomically thin stripes in transition metal dichalcogenide monolayers. *Nano Lett.* **16**, 6982–6987 (2016).
- Fei, L. et al. Direct TEM observations of growth mechanisms of two-dimensional MoS₂ flakes. *Nat. Commun.* **7**, 12206 (2016).
- Wu, H.-M. & Chen, S.-A. Dopant-polymer interaction: WCl₆ doped polyacetylene. *Synth. Met.* **20**, 169–183 (1987).
- Alov, N. V. XPS study of MoO₃ and WO₃ oxide surface modification by low-energy Ar⁺ ion bombardment. *Phys. Stat. Solidi C* **12**, 263–266 (2015).
- McGuire, G. E., Schweitz, G. K. & Carlson, T. A. Study of core electron binding-energies in some group IIIA, VB, and VIB compounds. *Inorg. Chem.* **12**, 2450–2453 (1973).

Acknowledgements This work was supported by the Singapore National Research Foundation under NRF award number NRF-NRFF2013-08, Tier 2 MOE2016-T2-2-153, MOE2016-T2-1-131, MOE2015-T2-2-007, Tier 1 RG164/15, Tier 1 RG4/17, CoE Industry Collaboration Grant WINTECH-NTU and the A*Star QTE programme. T.Y. acknowledges MOE Tier 1 RG100/15. J.L. and K.S. acknowledge the financial support of JST-ACCEL and JSPS KAKENHI (JP16H06333 and P16382). The work in SICCAS was supported by the National Key Research and Development Program of China (2016YFB0700204) and the National Natural Science Foundation of China (51502327). The work at IOP was supported by the Ministry of Science and Technology of China (grant numbers 2014CB920904, 2015CB921101 and 2016YFA0300600), the National Natural Science Foundation of China (grant numbers 11174340, 912212012, 11527806 and 91421303) and the Chinese Academy of Sciences (grant numbers XDB07010100). H.L. acknowledges the Singapore National Research Foundation for support under NRF award number NRF-NRFF2013-03. The work at Rice was supported by the US Department of Energy (DE-SC0012547) and by the R. Welch Foundation (C-1590).

Reviewer information Nature thanks D. Akinwande, M. Terrones and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.Z. and Z.L. designed the experiments. J.Z. grew all materials except TiSe₂, NbSe₂ and TaSe₂. H.W. grew TiSe₂, NbSe₂ and TaSe₂. Y.C., J.X., J.Z., T.Y. and Z.S. carried out Raman characterizations. J.L. and K.S. performed the STEM characterizations of all samples and analysis. J.Z. performed the AFM characterization of the samples. Y.Z., H.Y. and Q.L. performed the TG-DSC and XPS testing. Y.X., J.L. and B.I.Y. worked on ab initio reaction dynamics. C.-H.H., D.W. and H.L. performed electronic structure calculations. X.H., Q.Z., F.L. and Q.F. fabricated the devices. X.H., G.L., C.Y. and L.L. measured the superconductivity in MoTe₂ and NbSe₂. J.Z., J.L. and Z.L. wrote the paper. All authors discussed and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0008-3>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.L. or G.L. or Z.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

The 2D compounds and heterostructures were synthesized in a quartz tube with diameter 1 inch. The length of the furnace is about 36 cm. The system of the reaction is shown in Supplementary Fig. 32. A mixture of H_2/Ar was used as the carrier gas. Specifically, an aluminium oxide boat with volume about $8\text{ cm} \times 1.1\text{ cm} \times 1.2\text{ cm}$ containing precursor powder was placed in the centre of the tube. The precursor powder and the salt were mixed together first. The Si substrate with a 285 nm top layer of SiO_2 was placed on the aluminium oxide boat with the polished surface down; the distance between the sources and substrate ranges from 0.2 cm to 1.2 cm. Another aluminium boat containing S or Se or Te powder was placed on the upstream (upwind) of the tube furnace at 200 °C, 300 °C and 450 °C, respectively. The distance between the S or Se or Te boat and the precursor boat is about 18 cm, 16 cm and 15 cm, respectively. The heating rate of all reactions is 50 °C min^{-1} . All the reactions were carried out at atmospheric pressure. The temperature was cooled down to room temperature naturally. All reaction materials were bought from Alfa Aesar with purity more than 99%.

MoS₂. A powder mixture of 0.5 mg NaCl and 3 mg MoO_3 in an aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated to the growing temperature (600–800 °C) with a ramp rate of 50 °C min^{-1} . The growth time is 3–5 min. Ar (or Ar/H_2) with a flow rate of 80 (or 80/5) sccm (cubic centimetres per minute) was used as the carrier gas.

MoSe₂. The synthesis recipe for $MoSe_2$ is similar to that of MoS_2 . The growth temperature was about 700–800 °C, and Ar/H_2 with a flow rate of 80/5 sccm was used as the carrier gas.

MoTe₂. A powder mixture of 4 mg NaCl and 14 mg MoO_3 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated to the growth temperature (600–800 °C) with a ramp rate of 50 °C min^{-1} and held for 3 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 80/20 sccm was used as the carrier gas.

WS₂. Using our method, large-size and monolayered WS_2 single crystals can be prepared at a relatively low temperature of about 750–850 °C. An aluminium oxide boat containing 6 mg NaCl and 30 mg WO_3 was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–850 °C) and held for 3 min. Ar/H_2 at flow rate 80/10 sccm was used as the carrier gas.

WSe₂. The growth process of WSe_2 was similar to that of WS_2 . The size of the monolayered WSe_2 single crystal can reach 1 mm.

WTe₂. A powder mixture of 15 mg NaCl and 60 mg WO_3 in the boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–850 °C) and held at this temperature for 3 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 80/20 sccm was used as the carrier gas.

TiS₂. A powder mixture of 3 mg NaCl and 10 mg TiO_2 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–810 °C) and held at this temperature for 8–15 min before cooled down to room temperature naturally. Ar/H_2 with a flow rate of 80/20 sccm was used as the carrier gas. Using our method, monolayer TiS_2 with a size up to $50\text{ }\mu\text{m}$ was obtained.

TiSe₂. The parameters for the growth of $TiSe_2$ are similar to those of TiS_2 but replacing S with Se. Ar/H_2 with a flow rate of 100/20 sccm was used as the carrier gas.

TiTe₂. A powder mixture of 3 mg NaCl and 10 mg TiO_2 (note that Ti powder can be used) in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (800–850 °C) and held at this temperature for 10–15 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 110/20 sccm was used as the carrier gas.

ZrS₂. The synthesis procedures are as follows. A powder mixture of 3 mg NaCl and 10 mg ZrO_2 (note that Zr powder can be used) in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–800 °C) and held at this temperature for 10–15 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/20 sccm was used as the carrier gas.

ZrSe₂. Similar to the synthesis of ZrS_2 but replacing S with Se powder and raising the growth temperature to 750–830 °C.

ZrTe₂. Similar to the synthesis of ZrS_2 but replacing S with Te powder and raising the growth temperature to 800–850 °C.

HfS₂. The synthetic procedures are as follows. A powder mixture of 3 mg NaCl and 5 mg Hf in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (800–850 °C) and held at this temperature for 10–15 min before cooling down to room temperature naturally. Ar with a flow rate of 120 sccm was used as the carrier gas.

HfSe₂. Similar to the synthesis of HfS_2 but replacing S with Se powder and raising the growth temperature to 750–850 °C. Ar/H_2 with a flow rate of 120/20 sccm was used as the carrier gas.

HfTe₂. Similar to the synthesis of HfS_2 but replacing S with Te powder and raising the growth temperature to 800–850 °C. Ar/H_2 with a flow rate of 120/20 sccm was used as the carrier gas.

VS₂, VSe₂ and VTe₂. A powder mixture of 1 mg KI and 3 mg V_2O_5 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S/Se/Te powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (680–750 °C) and held at this temperature for 10–15 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 80/16 sccm was used as the carrier gas.

NbS₂, NbSe₂ and NbTe₂. A mixed powder of 2 mg NaCl and 10 mg Nb_2O_5 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S/Se/Te powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperatures (750–850 °C) and held at this temperature for 3–5 min before cooled down to room temperature naturally. Ar/H_2 with a flow rate of 80/16 sccm was used as carrier gas.

TaS₂, TaSe₂ and TaTe₂. A mixed powder of 5 mg NaCl and 30 mg Ta_2O_5 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S/Se/Te powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (800–850 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. The Ar/H_2 with a flow rate of 100/18 sccm was used as carrier gas.

ReS₂. A mixed powder of 1 mg KI and 5 mg Re in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (650–750 °C) and held for 5–10 min before cooling down to room temperature naturally. Ar with a flow rate of 60 sccm was used as the carrier gas.

ReSe₂. Although the synthesis of ReS_2 has been reported, monolayer $ReSe_2$ has not been synthesized successfully. Here, the preparation method of $ReSe_2$ is similar to the synthesis of ReS_2 . Se powder was used as the Se source and the reaction temperature was fixed at 700–780 °C. Ar/H_2 with a flow rate of 80/10 sccm was used as the carrier gas.

FeSe. A mixed powder of 2 mg NaCl and 10 mg Fe_2O_3 (or $FeCl_2$) in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing Se powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–850 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/18 sccm was used as the carrier gas.

PtS₂, PtSe₂, PtTe₂, PdS₂ and PdSe₂. A mixture of 1 mg NaCl and 10 mg M (using Pt, Pd nanoparticles or $PtCl_2$, $PdCl_2$ powder) in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S/Se/Te powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (800–850 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/20 sccm was used as the carrier gas.

MoS_{2-x}Te_{2(1-x)}. A powder mixture of 2 mg NaCl and 10 mg MoO_3 in the aluminium oxide boat was placed in the centre of the tube. Another aluminium oxide boat containing a powder mixture of S and Te was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature 700–800 °C and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/5 sccm was used as the carrier gas.

MoSe_{2-x}Te_{2(1-x)}. Similar to the synthesis of $MoS_{2-x}Te_{2(1-x)}$ except for using a powder mixture of Se and Te as precursor.

WS_{2-x}Te_{2(1-x)}. A powder mixture of 3 mg NaCl and 15 mg WO_3 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing a powder mixture of S and Te was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (750–850 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/5 sccm was used as the carrier gas.

WSe_{2-x}Te_{2(1-x)}. Similar to the synthesis of $WS_{2-x}Te_{2(1-x)}$ except for using a powder mixture of Se and Te as precursor.

NbS_{2-x}Se_{2(1-x)}. A powder mixture of 2 mg NaCl and 10 mg Nb_2O_5 in the aluminium oxide boat was placed in the centre of the tube. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (760–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H_2 with a flow rate of 100/15 sccm was used as the carrier gas.

Mo_{1-x}Nb_xSe₂. A powder mixture of 2 mg NaCl and 10 mg of $Nb_2O_5:MoO_3 = 1:1$ in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing Se powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min^{-1} to the growth temperature (760–840 °C) and held at this temperature for 10–20 min followed by cooling down

to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

Mo_{1-x}Re_xS₂. A powder mixture of 2 mg NaCl and 10 mg of Re:MoO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (700–800 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 80/5 sccm was used as the carrier gas.

W_{1-x}Nb_xS₂. A powder mixture of 2 mg NaCl and 15 mg of Nb₂O₅:WO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (750–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

W_{1-x}Nb_xSe₂. A powder mixture of 2 mg NaCl and 15 mg of Nb₂O₅:WO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (750–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

Mo_xNb_{1-x}S₂. A powder mixture of 2 mg NaCl and 10 mg of Nb₂O₅:MoO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (760–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

Mo_xW_{1-x}Te₂. A powder mixture of 2 mg NaCl and 10 mg of WO₃:MoO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (760–840 °C) and held at this temperature for 5–10 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

Mo_xNb_{1-x}S_{2-y}Se_{2(1-y)}. A powder mixture of 2 mg NaCl and 10 mg of Nb₂O₅:MoO₃ = 1:1 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S and Se powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (760–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/15 sccm was used as the carrier gas.

V_xW_yMo_{1-x-y}S_{2z}Se_{2(1-z)}. A powder mixture of 2 mg NaCl and 10 mg of V₂O₅:MoO₃:WO₃ = 1:5:3 in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing S and Se powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature (760–840 °C) and held at this temperature for 10–20 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 100/5 sccm was used as the carrier gas.

1 T' MoTe₂–2 H MoTe₂ in-plane heterostructures. A mixed powder of 4 mg NaCl and 14 mg MoO₃ in the aluminium oxide boat was placed in the centre of the quartz tube. The furnace was heated to a growth temperature of 720 °C with a ramp rate of 50 °C min⁻¹ and held for 3 min, and then quickly cooled to a growth temperature of 650 °C and held for 5 min and then cooled down to room temperature naturally. Ar/H₂ with flow rates of 80/20 sccm and 20/4 sccm was used as the carrier gas for 1 T' MoTe₂ and 2 H MoTe₂ growth, respectively.

MoS₂–NbSe₂ vertically stacked heterostructure. MoS₂ was synthesized first. A mixed powder of 0.5 mg NaCl and 3 mg MoO₃ in the aluminium oxide boat was placed in the centre of the tube. The furnace was heated to the growing temperature (600–800 °C) with a ramp rate of 50 °C min⁻¹. The growth time is 3 min. Ar (or Ar/H₂) with a flow rate of 80 sccm (or 80/5 sccm) was used as the carrier gas. The as-obtained MoS₂ was quickly transferred to another furnace for heterostructure growth. For the NbSe₂ growth, a mixed powder of 2 mg NaCl and 10 mg Nb₂O₅ and in the aluminium oxide boat was placed in the centre of the quartz tube. Another aluminium oxide boat containing Se powder was placed in the upstream. The furnace was heated with a ramp rate of 50 °C min⁻¹ to the growth temperature of 700 °C and held at this temperature for 10 min before cooling down to room temperature naturally. Ar/H₂ with a flow rate of 60/4 sccm was used as carrier gas.

We note that the weight ratio between salt and metal precursors can be tuned. Typically, for the synthesis of alloys, we fixed the weight ratio at 1:1 for the precursors. Taking Mo_{1-x}Re_xS₂ and WS_xTe_{2-x} as examples, by tuning the ratio between Mo and Re, and between S and Te, we can control the value of *x*.

STEM. STEM samples were prepared with a poly (methyl methacrylate) (PMMA)-assisted method or PMMA-free method with the assistance of an iso-propyl alcohol droplet. For some water-sensitive materials, we used a non-aqueous transfer method. STEM imaging and EELS analysis were performed on a JEOL 2100 F with a cold field-emission gun and an aberration corrector (the DELTA corrector) operating at 60 kV. A low-voltage modified Gatan GIF Quantum spectrometer was used for recording the EELS spectra. The inner and outer collection angles for the STEM image (β₁ and β₂) were 62 mrad and 129–140 mrad, respectively, with a convergence semi-angle of 35 mrad. The beam current was about 15 pA for the annular dark-field imaging and the EELS chemical analyses.

TG-DSC. TG-DSC measurements were performed using a Netzsch STA 449 C thermal analyser. Approximately 10 mg of the sample were loaded into an aluminium oxide crucible and heated at 10 K min⁻¹ from 20 °C to 920 °C. The 95 vol% Ar/5 vol% H₂ with a flow rate of 40 ml min⁻¹ was used as the carrier gas.

XPS. XPS measurements were performed using a monochromated Al Kα source (*hν* = 1486.6 eV) and a 128-channel mode detection Physical Electronics Inc. original detector. XPS spectra were acquired at a pass energy of 140 eV and a take-off angle of 45°.

Data availability. The main data supporting the findings of this study are available within the paper and its Supplementary Information. Extra data are available from the corresponding authors upon request.

Amino-acid- and peptide-directed synthesis of chiral plasmonic gold nanoparticles

Hye-Eun Lee^{1,6}, Hyo-Yong Ahn^{1,6}, Jungho Mun², Yoon Young Lee¹, Minkyung Kim³, Nam Heon Cho¹, Kiseok Chang⁴, Wook Sung Kim^{4,5}, Junsuk Rho^{2,3*} & Ki Tae Nam^{1*}

Understanding chirality, or handedness, in molecules is important because of the enantioselectivity that is observed in many biochemical reactions¹, and because of the recent development of chiral metamaterials with exceptional light-manipulating capabilities, such as polarization control^{2–4}, a negative refractive index⁵ and chiral sensing⁶. Chiral nanostructures have been produced using nanofabrication techniques such as lithography⁷ and molecular self-assembly^{8–11}, but large-scale and simple fabrication methods for three-dimensional chiral structures remain a challenge. In this regard, chirality transfer represents a simpler and more efficient method for controlling chiral morphology^{12–18}. Although a few studies^{18,19} have described the transfer of molecular chirality into micrometre-sized helical ceramic crystals, this technique has yet to be implemented for metal nanoparticles with sizes of hundreds of nanometres. Here we develop a strategy for synthesizing chiral gold nanoparticles that involves using amino acids and peptides to control the optical activity, handedness and chiral plasmonic resonance of the nanoparticles. The key requirement for achieving such chiral structures is the formation of high-Miller-index surfaces ($\{hkl\}$, $h \neq k \neq l \neq 0$) that are intrinsically chiral, owing to the presence of ‘kink’ sites^{20–22} in the nanoparticles during growth. The presence of chiral components at the inorganic surface of the nanoparticles and in the amino acids and peptides results in enantioselective interactions at the interface between these elements; these interactions lead to asymmetric evolution of the nanoparticles and the formation of helicoid morphologies that consist of highly twisted chiral elements. The gold nanoparticles that we grow display strong chiral plasmonic optical activity (a dissymmetry factor of 0.2), even when dispersed randomly in solution; this observation is supported by theoretical calculations and direct visualizations of macroscopic colour transformations. We anticipate that our strategy will aid in the rational design and fabrication of three-dimensional chiral nanostructures for use in plasmonic metamaterial applications.

To control the chiral morphology of gold nanoparticles through molecular interactions of amino acids or peptides with high-index surfaces, we devised an aqueous-based, two-step growth method involving organothiols additives. As the first step, low-index-plane-exposed gold nanoparticles of uniform size were synthesized using the well-established seed-mediated method^{23–25}. In the second step, cysteine or cysteine-based peptides with chiral conformations were used to encode chirality into the gold nanoparticles. The molecules were added to the growth solution, in which the pre-synthesized low-index-plane-exposed gold nanoparticles evolved into high-index-plane-exposed nanoparticles as a result of the reduction of Au³⁺ ions (see Methods for detailed experimental procedure). Au–S bonding and interactions of other functional groups in the amino acid or peptides are also involved in the nanoparticle growth process. Peptide-sequence-specific interactions have been investigated as a way of controlling the growth of nanomaterials and their optical properties^{26–28}. Changes in growth

components, such as the peptide sequence and concentration, and in seed morphology affect the growth kinetics and induce the dynamic morphological evolution of low-index-plane-exposed gold seed nanoparticles into chiral nanoparticles.

Circular dichroism and scanning electron microscopy (SEM) analyses confirm the synthesis of chiral plasmonic nanoparticles (Fig. 1). Notably, the conformation of the molecule used for synthesis controlled the handedness of the resulting nanoparticles. When L- or D- amino acids were added during the nanoparticle growth process, the nanoparticles that formed had the opposite handedness. Optical responses followed the handedness of nanoparticles. For example, when L-cysteine (L-Cys) and D-cysteine (D-Cys) were used as additives, the associated extinction spectra of the synthesized nanoparticles were identical and depended on only the overall particle size (Extended Data Fig. 1a). However, the measured circular dichroism spectra were inverted with respect to each other, but had the same peak positions, at 569 nm and 699 nm (Fig. 1a). In both the cases, the morphologies of the synthesized nanoparticles were cube-like, with a side length of 150 nm. An interesting feature of the nanoparticles synthesized using L-Cys and D-Cys is that the vertices protrude and the edges, which typically bridge two vertices in a cube, are split into two. As shown in the insets of Fig. 1b, the two split edges for the L-Cys nanoparticles point in opposite directions (one into and the other out of the cube), with a tilt angle of $-\varphi$. In the case of the D-Cys nanoparticles, the split edges were tilted in the opposite direction, at an angle of $+\varphi$ (Fig. 1c). Along the [111] view, the tilted edges protrude as tripods at each vertex of a cube, thereby contributing to the chirality of the synthesized gold nanoparticles (Fig. 1b, right inset). These tripods, which are 40 nm thick and 100 nm long, assemble, making nanometre-scale gaps inside a helicoid cube. The right-handed chiral structures, synthesized using L-Cys as an additive, exhibit increased absorption of left circularly polarized light at 569 nm, whereas the opposite chiro-optical response is observed for the left-handed chiral structures, synthesized using D-Cys. The yield of the chiral nanoparticles using this synthesis approach was about 81% (of a total of 989 nanoparticles) (Extended Data Fig. 1a).

The development of chiral morphology is a result of the different growth rates of the two oppositely chiral high-index planes of gold in the presence of L-Cys or D-Cys. Under our growth conditions, the absence of cysteine resulted in a stellated octahedron, differentiated by {321} facets (Fig. 2a, Extended Data Fig. 2a). (The synthesis method reported here can also be used to modify other stellated nanostructures^{29,30}.) We assigned the {321} indexing by analysing the relative angles of each edge in transmission electron microscopy (TEM) images (Extended Data Fig. 2c, d). The stellated octahedron has $4/m\bar{3}2/m$ point-group symmetry, defined by 48 identical triangular faces. The {321} facets are in the R (clockwise rotation, (321)^R) or S (anticlockwise rotation, (321)^S) conformation, defined by the rotational direction of the low-index planes (or microfacets) (100), (110) and (111), as outlined black in Fig. 2b^{20–22}. In Fig. 2a, the pairs of {hkl} planes with R and S conformation (within the rhombus ABA'B') are indicated in purple

¹Department of Materials Science and Engineering, Seoul National University, Seoul, South Korea. ²Department of Chemical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea. ³Department of Mechanical Engineering, POSTECH, Pohang, South Korea. ⁴R&D Center, LG Display, LG Science Park, Seoul, South Korea. ⁵Department of Electrical Engineering, POSTECH, Pohang, South Korea. ⁶These authors contributed equally: Hye-Eun Lee, Hyo-Yong Ahn. *e-mail: jsrho@postech.ac.kr; nkitae@snu.ac.kr

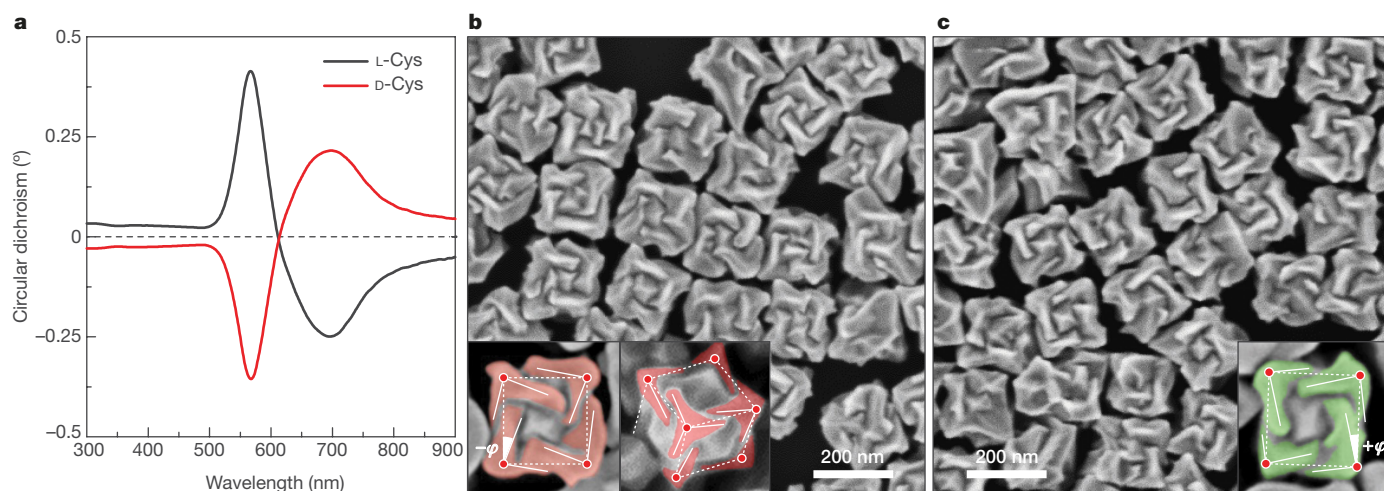


Fig. 1 | Opposite handedness of three-dimensional plasmonic helicoids controlled by cysteine chirality transfer. **a**, Circular dichroism spectra of chiral nanoparticles synthesized using L-Cys (black) and D-Cys (red). **b**, SEM image of L-Cys nanoparticles. The highlighting in the insets illustrates the fact that the edges (solid lines) are tilted by an angle $-\varphi$

with respect to the vertices (red dots) and cubic outline (dashed lines), as viewed along the [100] (left) and [111] (right) directions. **c**, SEM image of D-Cys nanoparticles. The inset highlights the tilted edges (solid lines), cubic outline (dashed lines) and tilt angle ($+\varphi$).

and yellow, respectively. The *R* and *S* triangular regions alternate and their distribution is essentially symmetric and achiral. We observe that the chiral morphologies developed as a result of the shifting and tilting of specific *R*–*S* boundaries. We present a detailed analysis of the time-dependent evolution when using L-Cys as the additive in the following.

The addition of low-index-plane-exposed cube-shaped seeds into the second-step solution containing L-Cys begins the growth. To monitor the evolution of morphology, nanoparticles were grown for different reaction times from 10 min to 120 min. The underlying mechanism of the evolution is most clearly evident in the 20-min case, in which the *g*-factor (see Methods) starts to increase over the next 20 min (Fig. 2c, d, Extended Data Fig. 4d). For clear visualization, the rhombus $\overline{ABA'B'}$, which consists of two sets of *R* and *S* regions, is displayed schematically (Fig. 2c, d, top) and marked with red dots and dotted white lines in the corresponding SEM images (Fig. 2c, d, bottom). Substantial changes, such as splitting, movement and overgrowth, were found at the *R*–*S* boundaries \overline{AC} and $\overline{A'C}$ in the rhombus $\overline{ABA'B'}$, and the 12 equivalent boundaries changed in the same manner. \overline{AC} and $\overline{A'C}$ were both tilted by $-\varphi$ towards the *S* regions and protruded with distortion, as indicated by the red-patterned area and arrows in Fig. 2c, d. In the [111] and [100] directions, the chiral elements formed three- and four-fold symmetry, respectively. As shown in the sequential images of the growth process (Extended Data Fig. 2e, f), the twisted edges continued to thicken, growing laterally and evolving into the final morphology, in which the elongated edges are twisted inwards (Fig. 2e). As the mirror symmetry of the *R* and *S* regions was broken by the distortion, the $4/m\bar{3}2/m$ point-group symmetry of the stellated octahedron changed to 432 symmetry. We therefore designate any nanoparticle with this chiral morphology as a ‘432 helicoid I’. When using D-Cys as the additive in the second-step solution, the *R*–*S* boundaries \overline{AC} and $\overline{A'C}$ are tilted by $+\varphi$ towards the *R* regions, resulting in a 432 helicoid I with opposite chirality.

In high-resolution TEM images (Extended Data Fig. 3a), the steps and terraces on the facets of a chiral nanoparticle at an early stage of growth (20 min) reveal the Cys adsorption on the high-index plane. In addition, the increased adsorption energy, as demonstrated by temperature-programmed desorption and electrochemical desorption studies, suggests that the molecules bind to a high-index surface (Extended Data Fig. 3b, c). N-terminal blocking of L-Cys completely inhibits the formation of chiral particles, and C-terminal blocking reduces the *g*-factor (Extended Data Fig. 5a). These data imply that

the thiol and amine groups bind with the ‘kinks’ on the {321} facets. This mechanism is also supported by previous studies^{31–33}, which have shown that the relative location of an amine group with respect to a thiol is the main determinant of the different binding affinities to *R* or *S* kink sites. Therefore, the preferred interaction of L-Cys with the {321} planes in the *R* regions leads to slower growth in the vertical direction on the *R* regions than on the *S* regions. For this reason, the *R*–*S* boundary shifts from the *R* to the *S* region, accompanied by asymmetric overgrowth.

The surface coverage of L-Cys is estimated to be 0.01 monolayers ($0.028 \text{ nmol cm}^{-2}$) at the optimal concentration (at which 432 helicoid I exhibits the highest *g*-factor) (Extended Data Fig. 4a–c, Methods). This low coverage seems to be necessary for chiral-selective growth—weak-binding motifs such as amine and carboxylic groups would be interfered with at high concentrations (Extended Data Fig. 4g). In addition, from a screening experiment with several peptide sequences, we found that other functional groups as well as thiol are key to determining the chiral morphology (Methods).

One of the most interesting results of our study is that the addition of L-glutathione (L-GSH) instead of L-Cys or D-Cys induces a completely different chiral morphology, by shifting a different *R*–*S* boundary (Fig. 2f–h). We observe a change in the four outer boundaries of the rhombus $\overline{ABA'B'}$ instead of in the inner \overline{AC} and $\overline{A'C}$ boundaries, as was observed in the case of L-Cys. Note that \overline{AC} and $\overline{A'C}$ are convex, whereas \overline{BC} and $\overline{B'C}$ are concave. \overline{AB} and $\overline{A'B'}$ both expand outwards, while the other boundaries ($\overline{AB'}$ and $\overline{A'B}$) move inwards, distorting the boundary of the rhombus $\overline{ABA'B'}$ (Fig. 2f). Consequently, a pinwheel-like chiral structure with clockwise rotation and four-fold symmetry appears along the [100] direction (Fig. 2g, h). We refer to any nanoparticle with this helicoid morphology as a ‘432 helicoid II’. A low-magnification SEM image shows uniformly synthesized 432 helicoid II particles (Extended Data Fig. 1b).

The different growth directions of 432 helicoids I and II can be understood at the atomic level by looking at the facets with $(321)^R$ conformation surrounded by those with $(312)^S$ and $(231)^S$ conformation in the [111] direction (Extended Data Fig. 5b–d). $(321)^R$ structures are composed of a (111) terrace and alternating (100) and (110) microfacets. Different orders of (100) and (110) alternation result in opposite chirality, such as $(312)^S$ or $(231)^S$. \overline{AC} , which is important in 432 helicoid I, is the boundary of $(321)^R$ and $(231)^S$; \overline{AB} , which is important in 432 helicoid II, is the boundary of $(321)^R$ and $(312)^S$. This property indicates that L-Cys and L-GSH shift the \overline{AC} boundary in the $[\bar{1}01]$

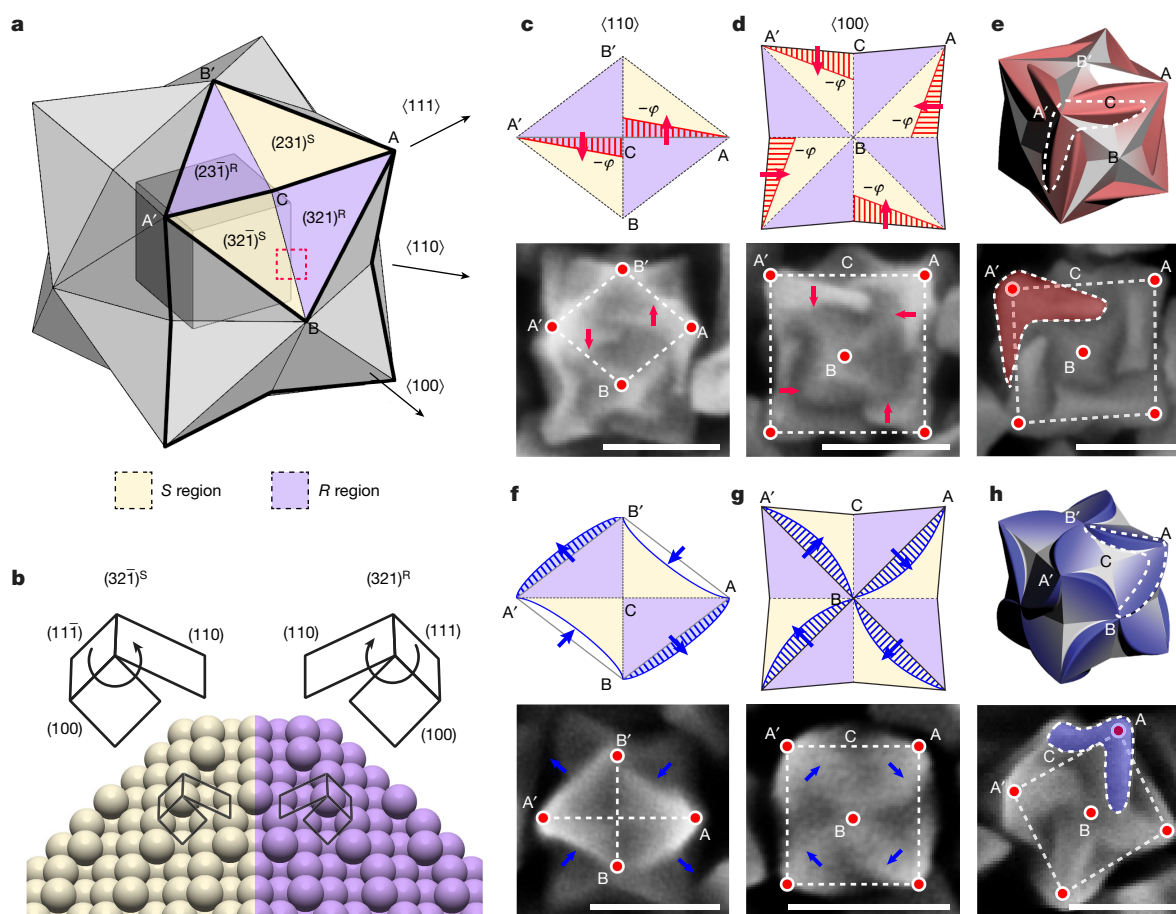


Fig. 2 | Mechanism of chirality evolution through the interplay between the enantioselective binding of molecules and the asymmetric growth of high-index facets. **a**, Schematic of a stellated octahedron, differentiated by high-index facets consisting of $\{321\}^S$ (S region, yellow) and $\{321\}^R$ (R region, purple) configurations. The vertices of the $\langle 111 \rangle$, $\langle 100 \rangle$ and $\langle 110 \rangle$ directions are indicated as A, B and C, respectively; A' and B' refer to the symmetric points of A and B, respectively. **b**, Comparison of the atomic arrangement of the $\{321\}^R$ and $\{321\}^S$ gold surfaces for region indicated by the red dotted box in **a**. The conformation at 'kink' sites is defined by the rotational direction of low-index microfacets in the sequence $(111) \rightarrow (100) \rightarrow (110)$: clockwise, R region; anticlockwise, S region. **c**, **d**, Schematics (top) and SEM images (bottom) of R–S pairs

showing the morphological development of 432 helicoid I in the presence of L-Cys, as viewed along the $[110]$ (**c**) and $[100]$ (**d**) directions. Newly developed boundaries are indicated as red patterned areas with arrows; each vertex is marked on the corresponding SEM image. **e**, Three-dimensional model (top) and SEM image (bottom) of the final chiral shape. The newly formed R region is coloured red and the chiral element is indicated by the dashed outline. The red shaded region in **e** is equivalent to the red patterned areas in **c** and **d**. **f**, **g**, Morphological development of 432 helicoid II in the presence of L-GSH. **h**, The corresponding final chiral shape. The newly formed boundaries and the pinwheel-like chiral elements of the final shape are coloured blue. Scale bars, 100 nm.

direction and the AB boundary in the $[01\bar{1}]$ direction, respectively. On the $\{321\}$ surfaces, the gold atoms attach to the (100) and (110) microfacets at the kink, generating a new kink. We propose that the orientation of the Cys or GSH molecule that is adsorbed could determine the specific growth direction of the kinks. Owing to the larger molecular size, GSH seems to interact with multiple kinks, whereas Cys interacts with only a single kink^{32,33}. We believe that the enantioselective interaction of L-GSH also benefits from the flexibility of the γ -peptide linkage, as supported by experiments with other GSH derivatives (Extended Data Fig. 5e). Calculations based on density functional theory and molecular dynamics are needed to identify other peptide sequences that affect single or multiple R–S boundaries and twist them multi-dimensionally.

The strongest optical activity among the nanoparticles that we synthesized was displayed by those that were synthesized using an octahedral seed instead of a cubic seed and therefore exhibit another type of chiral structure; we designate any such nanoparticle as a '432 helicoid III'. 432 helicoid III nanoparticles have pinwheel-like structures—consisting of four highly curved arms of increasing width—on each of the six faces of the cubic geometry (Fig. 3a, b, Extended Data Fig. 6a). Compared to 432 helicoids I and II, the chiral elements in 432 helicoid III are twisted with larger curvature and the gaps between them are

carved more deeply in the central direction. Imaging after ion milling using helium-ion microscopy shows the curved surfaces located inside the gaps (Extended Data Fig. 6b). From the depth and curvature information, we construct a three-dimensional model to assign the Miller index at each location (Extended Data Fig. 6c, d). The strong circular dichroism signal of this structure (Fig. 3c) is largely attributed to the highly twisted chiral structures and is consistent with the simulation results (Fig. 3d). The *g*-factor of 432 helicoid III is approximately 0.2 at 622 nm, which is roughly ten and three times larger than that of 432 helicoids I and II, respectively (Fig. 3e). The *g*-factors of various chiral nanostructures are compared in Extended Data Table 1. The *g*-factor of 432 helicoid III is larger than that of any other chiral nanostructure fabricated using bottom-up approaches. The exceptionally strong chiro-optical properties of 432 helicoid III may have resulted from the high-order plasmonic modes of large continuous chiral particles (Extended Data Fig. 7). The intensity difference for the electric and magnetic near-fields (Fig. 3f) is consistent with the macroscopic asymmetric response of 432 helicoid III to circularly polarized light. Interestingly, in contrast to a symmetric sphere, the induced magnetic dipole moment of 432 helicoid III cannot be perpendicular to the induced electric dipole moment. We also find that several other structural changes, such as edge length, gap width, gap depth, gap angle

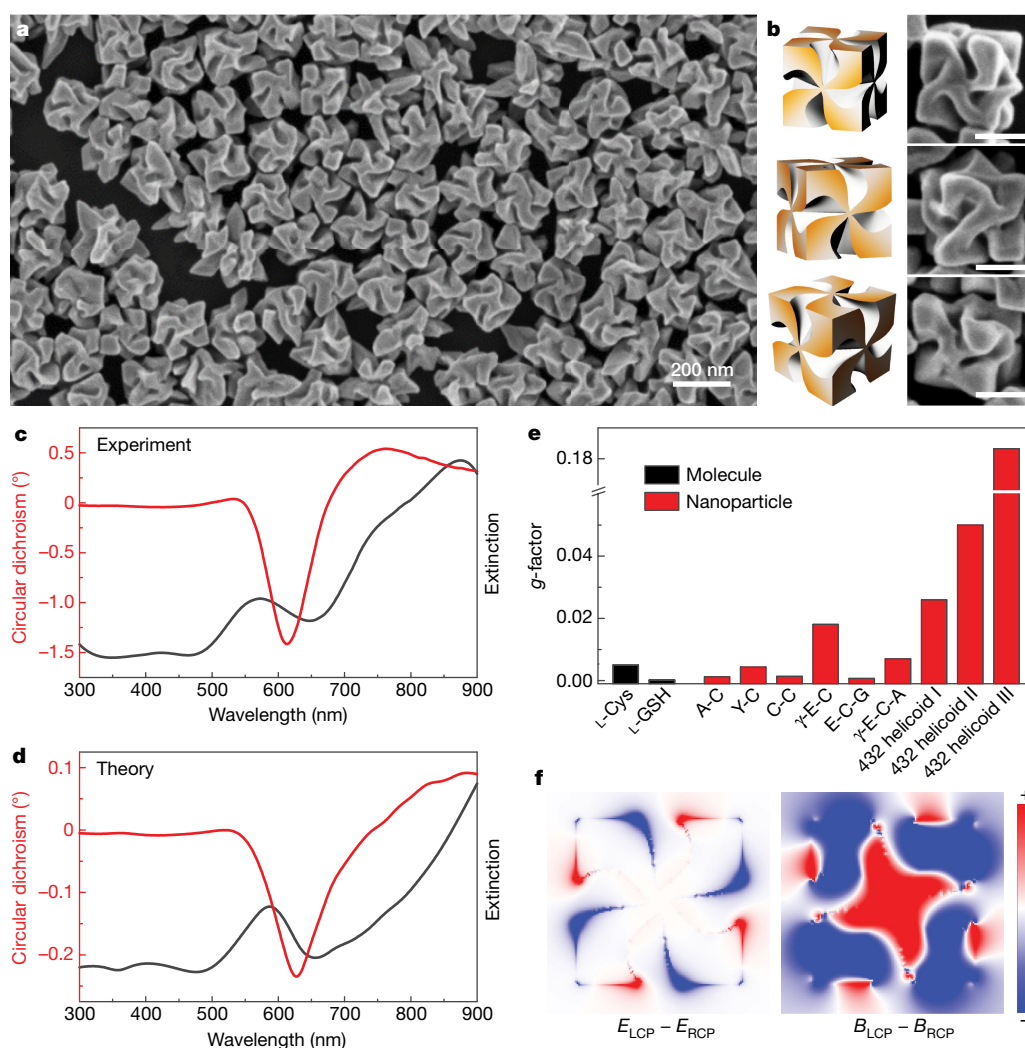


Fig. 3 | Morphology and optical activity of 432 helicoid III. **a**, SEM image of 432 helicoid III nanoparticles evolved from an octahedral seed. **b**, Three-dimensional models (left) and corresponding SEM images (right) of 432 helicoid III oriented in various directions. Scale bars, 100 nm. **c**, **d**, Experimental (**c**) and theoretical (**d**; based on a finite-difference time-domain method) circular dichroism and extinction spectra of 432 helicoid III. **e**, Comparison of the dis-symmetry *g*-factors of the synthesized helicoid structures and other nanoparticles (SEM images are shown in Extended Data Fig. 5e, f). A-C, L-alanyl-L-cysteine; Y-C,

L-tyrosyl-L-cysteine; C-C, L-cysteinyl-L-cysteine; γ-E-C, γ-L-glutamyl-L-cysteine; E-C-G, L-glutamyl-L-cysteinyl-glycine; γ-E-C-A, γ-L-glutamyl-L-cysteinyl-L-alanine. **f**, Theoretical calculation of the dependence of local electromagnetic fields on the handedness of circularly polarized light. The asymmetric responses of the electric (left; *E*) and magnetic (right; *B*) fields are displayed by the differences in these fields under excitation by left circularly polarized (LCP) and right circularly polarized (RCP) light. The colour scale indicates the magnitude of the field difference, with red (blue) indicating a positive (negative) difference.

and curvature, affect the chiro-optical activity of the nanoparticles (Extended Data Figs 8, 9, Methods).

As a result of the large *g*-factor of the helicoid nanoparticles, a macroscopically distinguishable change in their colour is possible by controlling the polarization. The circular dichroism spectrum and the corresponding optical rotatory dispersion spectrum for 432 helicoid III are presented in Fig. 4a, and the output polarization state was measured directly at four wavelengths using linearly polarized incident light (Fig. 4b). The largest ellipticity ($\chi = -28.7^\circ$, left circular polarization) was observed at 635 nm and the azimuthal rotation (ψ) changed gradually from -7.9° to $+29^\circ$ as the wavelength was increased. The conversion from linearly to elliptically polarized light by 432 helicoid III is clear under cross-polarized conditions. Although the achiral nanoparticles did not exhibit any transmission (Fig. 4c, left), a solution of 432 helicoid III nanoparticles showed bright yellow cross-polarized transmission, which reflects a pronounced polarization-rotating ability at visible wavelengths (Fig. 4c, right). Further, we confirmed the isotropic response and Lorentz reciprocity of the nanoparticles (Extended Data Fig. 10a, b). Changing the size of 432 helicoid III nanoparticles by controlling the initial seed concentration caused a resonance shift of the

resulting nanoparticles, with λ_{\max} (the wavelength at which the maximum *g*-factor is observed) increasing from 552 nm to 668 nm (Fig. 4d, Extended Data Fig. 10c). This modification also enabled gradual tuning of the transmitted colours under cross-polarized conditions (0° , dotted box in Fig. 4e). In addition, rotation of the analyser reversibly generated various transmitted colours, thereby providing a versatile method of colour modulation that reflects the optical rotatory dispersion response (Fig. 4e, Extended Data Fig. 10d, Supplementary Video 1). In contrast to the symmetric pattern for achiral nanoparticles, the colour transition of 432 helicoid III was continuous and asymmetric (Fig. 4e), forming elliptical traces in the chromaticity diagram (Extended Data Fig. 10e–h). The colour transformation of 432 helicoid III was dynamic and covered a wide range of colours.

We envision that the biomolecular approach presented here for the evolution of chirality in a plasmonic helicoid has technological potential for the development of biologically responsive and tunable metamaterials. Using this approach, chiral elements were arranged within cube-like structures with a side length of only about 100 nm, resulting in three-dimensional, angle-insensitive plasmonic metamaterials. We believe that conformation control using long peptides or other chiral

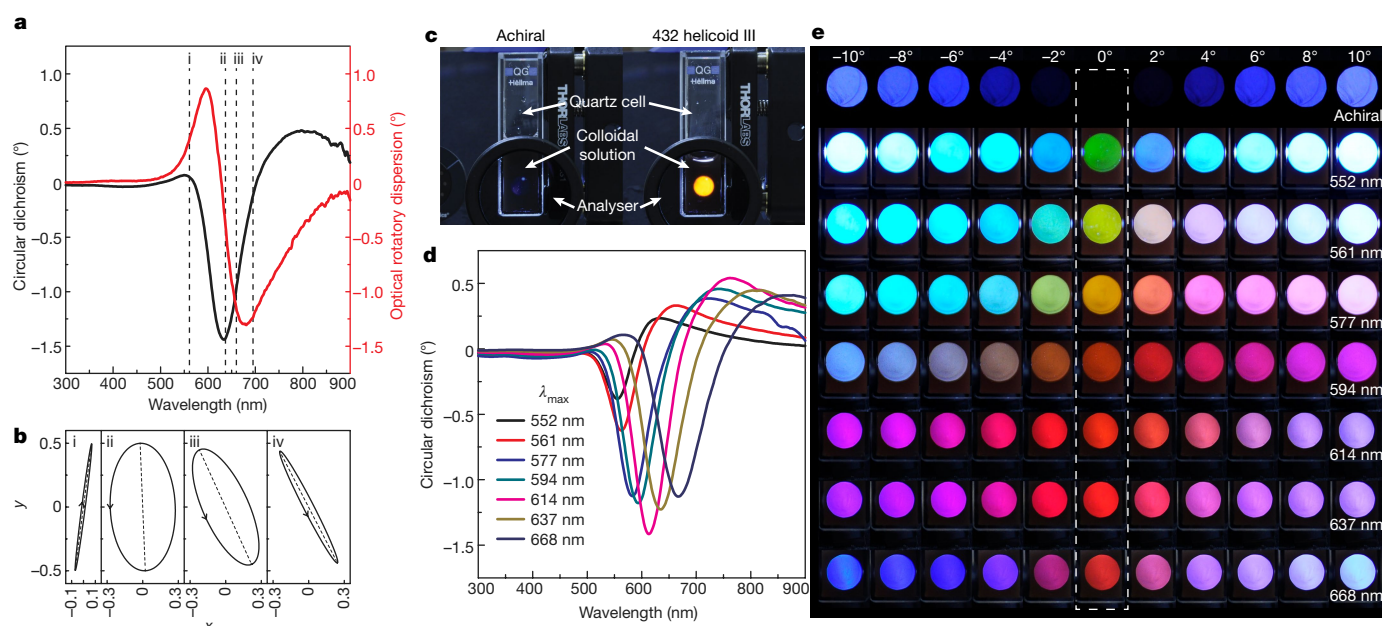


Fig. 4 | Visible light polarization control by 432 helicoid III solution. **a**, Circular dichroism and optical rotatory dispersion spectra of 432 helicoid III. **b**, Output polarization states measured at the wavelengths indicated in **a**: 561 nm (i), 635 nm (ii), 658 nm (iii) and 690 nm (iv). The polarization ellipses at each wavelength expressed in terms of the ellipticity (χ) and azimuthal rotation (ψ) are: $\chi = 1.7^\circ$ and $\psi = -7.9^\circ$ at 561 nm; $\chi = -28.7^\circ$ and $\psi = 2.6^\circ$ at 635 nm; $\chi = -20.7^\circ$ and $\psi = 26.6^\circ$ at 658 nm; and $\chi = -4.8^\circ$ and $\psi = 29.0^\circ$ at 690 nm (x and y are the horizontal and vertical components of the electric-field vector). **c**, Photographs of achiral (left) and 432 helicoid III (right) solutions, showing the light transmitted

under cross-polarized conditions. **d**, Circular dichroism spectra of 432 helicoid III for different values of λ_{\max} (the wavelength of the maximum g -factor). Seven spectra were obtained from differently sized 432 helicoid III nanoparticles with various resonance peaks (552 nm, 561 nm, 577 nm, 594 nm, 614 nm, 637 nm and 668 nm). **e**, Polarization-resolved colours of light transmitted through seven different 432 helicoid III solutions with different λ_{\max} values (shown on the right) and an achiral nanoparticle solution (top row). The rotational angle of the analyser was increased from -10° (left-most) to 10° (right-most) (see Methods). An angle of 0° , indicated by the dashed box, represents cross-polarized conditions.

biomolecules will enable the synthesis of other sets of chiral symmetry groups. Further, insights from this study could provide theoretical guidelines for designing artificial chirality and chiro-optical properties for active colour displays, holography, reconfigurable switching, chirality sensing and all-angle negative-refractive-index materials.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0034-1>.

Received: 29 May 2017; Accepted: 26 February 2018;

Published online 18 April 2018.

- Hazen, R. M. & Sholl, D. S. Chiral selection on inorganic crystalline surfaces. *Nat. Mater.* **2**, 367–374 (2003).
- Gansel, J. K. et al. Gold helix photonic metamaterial as broadband circular polarizer. *Science* **325**, 1513–1515 (2009).
- Meinzer, N., Barnes, W. L. & Hooper, I. R. Plasmonic meta-atoms and metasurfaces. *Nat. Photon.* **8**, 889–898 (2014).
- Zhang, S. et al. Photoinduced handedness switching in terahertz chiral metamolecules. *Nat. Commun.* **3**, 942 (2012).
- Pendry, J. B. A chiral route to negative refraction. *Science* **306**, 1353–1355 (2004).
- Hendry, E. et al. Ultrasensitive detection and characterization of biomolecules using superchiral fields. *Nat. Nanotechnol.* **5**, 783–787 (2010).
- Soukoulis, C. M. & Wegener, M. Past achievements and future challenges in the development of three-dimensional photonic metamaterials. *Nat. Photon.* **5**, 523–530 (2011).
- Ma, W. et al. Chiral inorganic nanostructures. *Chem. Rev.* **117**, 8041–8093 (2017).
- Kuzyk, A. et al. DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, 311–314 (2012).
- Kuzyk, A. et al. Reconfigurable 3D plasmonic metamolecules. *Nat. Mater.* **13**, 862–866 (2014).
- Lee, H.-E., Ahn, H.-Y., Lee, J. & Nam, K. T. Biomolecule-enabled chiral assembly of plasmonic nanostructures. *ChemNanoMat* **3**, 685–697 (2017).
- Kühnle, A., Linderth, T. R., Hammer, B. & Besenbacher, F. Chiral recognition in dimerization of adsorbed cysteine observed by scanning tunnelling microscopy. *Nature* **415**, 891–893 (2002).

- Switzer, J. A., Kothari, H. M., Poizot, P., Nakanishi, S. & Bohannon, E. W. Enantiospecific electrodeposition of a chiral catalyst. *Nature* **425**, 490–493 (2003).
- Horvath, J. D., Koritnik, A., Kamakoti, P., Sholl, D. S. & Gellman, A. J. Enantioselective separation on a naturally chiral surface. *J. Am. Chem. Soc.* **126**, 14988–14994 (2004).
- Xiao, W. et al. Microscopic origin of chiral shape induction in achiral crystals. *Nat. Chem.* **8**, 326–330 (2016).
- Sanchez, C., Arribart, H. & Guille, M. M. G. Biomimetic and bioinspiration as tools for the design of innovative materials and systems. *Nat. Mater.* **4**, 277–288 (2005).
- Weissbuch, I., Addadi, L., Lahav, M. & Leiserowitz, L. Molecular recognition at crystal interfaces. *Science* **253**, 637–645 (1991).
- Che, S. et al. Synthesis and characterization of chiral mesoporous silica. *Nature* **429**, 281–284 (2004).
- Orme, C. A. et al. Formation of chiral morphologies through selective binding of amino acids to calcite surface steps. *Nature* **411**, 775–779 (2001).
- McFadden, C. F., Cremer, P. S. & Gellman, A. J. Adsorption of chiral alcohols on “chiral” metal surfaces. *Langmuir* **12**, 2483–2487 (1996).
- Ahmadi, A., Attard, G., Feliu, J. & Rodes, A. Surface reactivity at “chiral” platinum surfaces. *Langmuir* **15**, 2420–2424 (1999).
- Sholl, D. S., Asthagiri, A. & Power, T. D. Naturally chiral metal surfaces as enantiospecific adsorbents. *J. Phys. Chem. B* **105**, 4771–4782 (2001).
- Sau, T. K. & Murphy, C. J. Room temperature, high-yield synthesis of multiple shapes of gold nanoparticles in aqueous solution. *J. Am. Chem. Soc.* **126**, 8648–8649 (2004).
- Ahn, H.-Y., Lee, H.-E., Jin, K. & Nam, K. T. Extended gold nano-morphology diagram: synthesis of rhombic dodecahedra using CTAB and ascorbic acid. *J. Mater. Chem. C* **1**, 6861–6868 (2013).
- Lee, H.-E. et al. Concave rhombic dodecahedral Au nanocatalyst with multiple high-index facets for CO₂ reduction. *ACS Nano* **9**, 8384–8393 (2015).
- Dickerson, M. B., Sandhage, K. H. & Naik, R. R. Protein- and peptide-directed syntheses of inorganic materials. *Chem. Rev.* **108**, 4935–4978 (2008).
- Slocik, J. M., Govorov, A. O. & Naik, R. R. Plasmonic circular dichroism of peptide-functionalized gold nanoparticles. *Nano Lett.* **11**, 701–705 (2011).
- Coppage, R. et al. Determining peptide sequence effects that control the size, structure, and function of nanoparticles. *ACS Nano* **6**, 1625–1636 (2012).
- Liu, H. et al. Stellated Ag-Pt bimetallic nanoparticles: an effective platform for catalytic activity tuning. *Sci. Rep.* **4**, 3969 (2014).
- Weiner, R. G. & Skrabalak, S. E. Metal dendrimers: synthesis of hierarchically stellated nanocrystals by sequential seed-directed overgrowth. *Angew. Chem. Int. Ed.* **54**, 1181–1184 (2015).

31. Kühnle, A., Linderoth, T. R. & Besenbacher, F. Enantiospecific adsorption of cysteine at chiral kink sites on Au(110)-(1×2). *J. Am. Chem. Soc.* **128**, 1076–1077 (2006).
32. Greber, T., Slijvančanin, Z., Schillinger, R., Wider, J. & Hammer, B. Chiral recognition of organic molecules by atomic kinks on surfaces. *Phys. Rev. Lett.* **96**, 056103 (2006).
33. Schillinger, R., Slijvančanin, Z., Hammer, B. & Greber, T. Probing enantioselectivity with x-ray photoelectron spectroscopy and density functional theory. *Phys. Rev. Lett.* **98**, 136102 (2007).

Acknowledgements This research was supported by a Seoul National University research grant in 2015, the LG Display under LGD-SNU Incubation Program, the Creative Materials Discovery Program (2017M3D1A1039377), which is funded by the National Research Foundation (NRF) under the Ministry of Science, ICT and Future Planning (MSIP), South Korea, and the KIST-SNU Joint Research Program (0543-20180021). K.T.N. acknowledges financial supports from the Global Frontier R&D Program of the Center for Multiscale Energy System (2012M3A6A7054855), which is funded by the NRF-MSIP, South Korea. J.R. acknowledges financial support from the Engineering Research Center Program of the Center for Optically-assisted Mechanical Systems (2015R1A5A1037668) and the Global Frontier R&D Program of the Center for Advanced Meta-Materials (2014M3A6B3063708), which is funded by the NRF-MSIP, South Korea. H.-Y.A. and M.K. are grateful to the Global PhD Fellowship Program (2014H1A2A1020809 and 2017H1A2A1043204, respectively), which is funded by the NRF-MSIP,

South Korea. J.R. thanks T. Liedl (Ludwig Maximilian University of Munich) and N. Liu (University of Heidelberg) for discussions.

Reviewer information *Nature* thanks L. M. Liz-Marzán and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.T.N. and J.R. conceived the idea. H.-E.L., H.-Y.A., N.H.C. and Y.Y.L. synthesized and characterized the materials. J.M., M.K. and J.R. performed the numerical simulations and analysed the data. H.-E.L., H.-Y.A., K.C. and W.S.K. conducted the optical characterization of the materials. All authors discussed the experiments and contributed to writing the manuscript. J.R. guided the numerical simulations and optical measurements. K.T.N. guided all aspects of the work.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0034-1>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0034-1>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.R. or K.T.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Chemicals. Hexadecyltrimethylammonium bromide (CTAB, 99%), L-ascorbic acid (AA, 99%) and tetrachloroauric(III) trihydrate ($\text{HAuCl}_4 \cdot 3\text{H}_2\text{O}$, 99.9%) were purchased from Sigma-Aldrich. L-cysteine hydrochloride monohydrate (99%, TCI), D-cysteine hydrochloride monohydrate (99%, TCI), L-cysteine ethyl ester hydrochloride (99%, TCI), N-acetyl-L-cysteine (98%, TCI), L-glutathione (γ -E-C-G, 98%, Sigma-Aldrich), L-glutathione ethyl ester (90%, Sigma-Aldrich) and γ -L-glutamyl-L-cysteine (γ -E-C, 80%, Sigma-Aldrich) were obtained commercially and used without further purification. Di- and tri-peptides, L-alanyl-L-cysteine (A-C, > 98%), L-prolyl-L-cysteine (P-C, > 98%), L-cysteinyl-L-cysteine (C-C, > 98%), L-tyrosyl-L-cysteine (Y-C, > 98%), L-glutamyl-L-cysteinylglycine (E-C-G, > 98%) and γ -L-glutamyl-L-cysteinyl-L-alanine (γ -E-C-A, > 98%) were provided by GenScript and prepared in hydrochloride salt form before use. All aqueous solutions were prepared using high-purity deionized water ($18.2 \text{ M}\Omega \text{ cm}^{-1}$).

Synthesis of chiral nanoparticles. Cubic and octahedral seeds were synthesized as reported previously^{24,34}. Before use, both types of seed nanoparticle were centrifuged (6,708g, 150 s) twice and dispersed in aqueous CTAB (1 mM) solution. In a typical synthesis, a growth solution was prepared by adding 0.8 ml of 100 mM CTAB and 0.2 ml of 10 mM gold chloride trihydrate into 3.95 ml of deionized water to form an $[\text{AuBr}_4]^-$ complex. Au^{3+} was then reduced to Au^0 by the rapid injection of 0.475 ml of 100 mM AA solution. The growth of chiral nanoparticles was initiated by adding 5 μl of amino acid or peptide solution and 50 μl of seed solution into the growth solution. For the preparation of 432 helicoid I, cubic seed solution was added to the growth solution and then, after a 20-min incubation, 100 μM cysteine was added. To prepare 432 helicoid II, 2 mM glutathione was added to the growth solution, followed by the addition of cubic seed solution. To prepare 432 helicoid III, 5 mM glutathione was added to the growth solution, followed by the addition of octahedral seed solution. The growth solution was placed in a 30 °C bath for 2 h, and the pink solution gradually became blue with large scattering. The solution was centrifuged twice (1,677g, 60 s) to remove unreacted reagents and was re-dispersed in a 1 mM CTAB solution for further characterization.

Characterization. Extinction and circular dichroism (CD) spectra were obtained using a J-815 spectropolarimeter instrument (JASCO), and optical rotatory dispersion (ORD) spectra were measured using an additional ORD attachment. To check the Lorentz reciprocity, we prepared solutions of nanoparticles and particles attached on the substrate. The CD spectrum of each sample condition was measured in the forwards and backwards directions by changing the direction of the sample relative to the incident light.

Kuhn's dis-symmetry factor (*g*-factor) is a dimensionless quantity that is useful for quantitative comparisons of chiro-optical properties among different systems and was calculated from the measured extinction and CD values using:

$$g\text{-factor} = 2 \frac{A_L - A_R}{A_L + A_R} \propto \frac{\text{CD}}{\text{extinction}}$$

SEM images were taken with a SIGMA system (Zeiss). TEM images were captured using a JEM-3000F system (JEOL).

The polarization-rotating ability of 432 helicoid III was evaluated from polarization-state measurements using an optical configuration consisting of a laser source, iris, linear polarizer, quarter-wave plate, sample and polarimeter. The output polarization state was measured using a PAX5710VIS-T rotating-wave plate Stokes polarimeter (Thorlabs). Laser sources with centre wavelengths of 561 nm (CNI MGL-FN-561, DPSS Laser), 635 nm (Hitachi HL6321G laser diode), 658 nm (Hitachi HL6501MG laser diode) and 690 nm (Hitachi HL6738MG laser diode) were used. For the measurements, a solution of randomly dispersed 432 helicoid III nanoparticles was added to a quartz cell with a path length of 10 mm, and was then irradiated with a vertically polarized incident beam. A quarter-wave plate was used with a linear polarizer to compensate for any polarization interference caused by other optical parts of the system.

Macroscopic colour changes in transmitted light were detected by polarization-resolved transmission measurements with an optical configuration consisting of a white-light illumination source, iris, linear polarizer, sample, linear polarizer (analyser) and digital camera. The sample was placed between two crossed linear polarizers (0° represents cross-polarized conditions) and was irradiated with a collimated cold white-light source. The angle of the analyser was changed from -10° (clockwise) to +10° (anticlockwise) in steps of 1° from the orthogonal configuration, which enables different wavelengths of light to propagate, rendering different colours of transmitted light. While rotating the analyser, the colour transition of the transmitted light was observable by the naked eye and was recorded with a digital camera (D90, Nikon).

Quantification of amino acids and peptides. To quantify the amount of Cys and GSH on the 432 helicoids I and II, a thiol-selective dye-based fluorometric assay was performed. After complete growth in growth solution (100 nM L-Cys

for 432 helicoid I and 2.5 μM L-GSH for 432 helicoid II), the chiral nanoparticles were centrifuged and washed three times to remove the remaining chemicals in solution except the molecules adsorbed on the nanoparticle surface. By adding NaBH_4 to the nanoparticle solution, the reductive desorption reaction of adsorbed thiolate molecules (Au-SR) started immediately, and free thiol molecules (RS^-) were released to the solution as follows: $\text{Au-SR} + e^- \rightarrow \text{Au} + \text{RS}^-$ (Extended Data Fig. 4a)^{35,36}. The final concentration of NaBH_4 was 25 mM and the final volume of the solution was fixed to 100 μl . After 5 min of incubation, the nanoparticles were centrifuged again and clear supernatant solutions containing the released Cys or GSH molecules were collected and incubated for 1 day in 25 °C to decompose the remaining NaBH_4 .

Quantification of Cys and GSH on the 432 helicoids I and II was carried out by using thiol-selective dye (Thiol detection assay kit, Cayman Chemical, 700340). The sample was diluted in the reaction buffer (10 mM phosphate buffer with 1 mM EDTA, pH 7.4). The thiol-selective dye reacted spontaneously with the free thiol of Cys or GSH in the sample solution, producing a fluorescent derivative (Extended Data Fig. 4a). The fluorescence signal of the sample was recorded by using an excitation wavelength of 405 nm and an emission wavelength of 535 nm. For the relative quantification, the standard solution was prepared under the same conditions and measured in the concentration range 0–5 μM . The standard concentration curve showed good linearity, with $R^2 = 0.999$ (Extended Data Fig. 4b). The quantified amounts of Cys and GSH on the surface of 432 helicoids I and II were $98.9 \pm 22.6 \text{ pmol}$ and $280.2 \pm 90.2 \text{ pmol}$, respectively.

The adsorption amount of GSH during the synthesis of 432 helicoid II was monitored over time (Extended Data Fig. 4e). To stop the growth at different stages, the particles were centrifuged out every 10 min. After the centrifugation was repeated three times to remove the remaining chemicals, the quantification experiment for GSH was performed as described above.

Calculation of surface coverage. To convert the measured concentration of molecules to the surface coverage, we estimated the total surface area of the 432 helicoid I and II samples. According to extinction measurements for the seed nanoparticles, the total number of nanoparticles in each sample batch was measured to be $N_{\text{NP}} = 1.53 \times 10^9 \text{ ml}^{-1}$. The surface area of a single 432 helicoid I nanoparticle was $A_{\text{NP,H1}} = 2.31 \times 10^{-9} \text{ cm}^2$, and that of helicoid II was $A_{\text{NP,H2}} = 1.78 \times 10^{-9} \text{ cm}^2$, approximated from the schematic three-dimensional models in Extended Data Fig. 1. Therefore, the total surface area of the nanoparticles in each sample were

$$A_{\text{total}} = N_{\text{NP}} A_{\text{NP}}, \quad A_{\text{total,H1}} = 3.53 \text{ cm}^2, \quad A_{\text{total,H2}} = 2.72 \text{ cm}^2$$

The surface coverage of Cys and GSH in 432 helicoids I and II was calculated from the quantification results and the estimated total surface area of the nanoparticle samples, as shown in Extended Data Fig. 4c. The surface density of Cys and GSH was 0.028 nmol cm^{-2} and 0.103 nmol cm^{-2} , respectively. These values correspond to 0.01 and 0.22 monolayers, respectively, calculated from the reported maximum coverage^{37,38}. On the basis of the surface density estimate, the average intermolecular distance for Cys and GSH is expected to be 2.5 nm and 1.3 nm, respectively. Therefore, we conclude that, under optimized conditions, both molecules have enough room to bind multiple functional groups, as is necessary for enantioselective recognition.

Adsorption study of Cys and GSH on {321} nanoparticles. To compare adsorption kinetics, we measured the amount of adsorbed Cys and GSH on {321} nanoparticles. The {321} nanoparticles were incubated in Cys and GSH solution with different concentrations from 0.5 μM to 10 μM . After 2 h of incubation, the nanoparticles were centrifuged and the clear supernatant solutions that contained the remaining molecules were collected for quantification using thiol-specific dye. The adsorbed amount was calculated by subtracting the measured supernatant concentration from the initial concentration (Extended Data Fig. 4f). Given the concentration, a larger amount of Cys was attached to the high-indexed surface compared to GSH. In addition, a similar trend is reflected in the formation of chiral morphology. Owing to this fast loading of Cys, only 0.1 μM is required for 432 helicoid I, whereas molecules that are 25 times larger are needed for 432 helicoid II to achieve a chiral morphology (Extended Data Fig. 4g, h).

Effects of functional group on chiral morphology. We investigated the effects of functional groups on the resulting morphology of the gold nanoparticles to further understand the amino acid and peptide interactions with the gold surface at a molecular level (Extended Data Fig. 5). The N terminus of the amino acid or peptide that was added was a critical parameter in determining the chiral shape. For example, blocking an amine group in L-Cys greatly reduced chirality (Extended Data Fig. 5a). Different N-terminal sequences modified the morphology and texture of the gold nanoparticle surface considerably. We examined several dipeptides with N-terminal modifications of alanine, proline, cysteine and tyrosine (Extended Data Fig. 5f). The morphology of the resulting nanoparticles was very dependent on the peptide sequence. Such substantial morphological differences may arise from alterations in the binding sites and the spatial arrangement of functional

groups that are imparted by dipeptide side chains. Compared to Cys, GSH has an elongated N terminus owing to the specific γ -Glu group. When γ -Glu is replaced with α -Glu (E-C-G; Extended Data Fig. 5e), the chiral morphology was noticeably degraded. In addition, nanoparticle synthesis using γ -E-C produced a different morphology with a certain level of chirality, whereas other dipeptide cases resulted in achiral morphologies (Extended Data Fig. 5f). These results support the idea that the γ -Glu group has an important role in the evolution of chirality.

In addition to the N-terminal modification, substitution of the C terminus resulted in different types of shape evolution owing to changes in the spatial arrangement of functional groups related to the oriented attachment of amino acids and peptides^{12,31}. When L-Cys ethyl ester with a blocked carboxylic acid group was used in the synthesis, the *g*-factor was reduced by a factor of approximately ten (Extended Data Fig. 5a). Blocking of the C-terminal carboxylic acid of L-GSH (γ -E-C-G) generated only achiral structures (Extended Data Fig. 5e). In addition, a different chiral morphology was developed by replacing Gly with Ala, probably as a result of the steric hindrance near the C-terminal side. According to a previous report³⁹, the -COOH of the Gly moiety in GSH is involved in binding onto the gold surface, along with the thiol and amine groups. The different chiral structure that is induced by the exchange of sequence at the C terminus suggests that more diversified chiral structures may be synthesized by changing the C-terminal sequence (Extended Data Fig. 5e).

Temporal evolution of chiral nanoparticles. The different molecular features of Cys and GSH collectively influenced the morphological development and thus led to notable changes in the final chiral morphology. To monitor the evolution of chirality, the growth reaction was stopped at different stages by centrifugation, after which we performed three repetitions of washing, re-dispersion and centrifugation to remove the remaining chemicals. To obtain a detailed comparison of chiral evolution, we analysed the temporal growth of 432 helicoids I and II in terms of SEM (Extended Data Fig. 2e, f), *g*-factor (Extended Data Fig. 4d) and the amount of GSH in a nanoparticle (Extended Data Fig. 4e).

In the case of 432 helicoid I, the AC and A'C boundaries between the R and S planes started to develop and shift slightly to the S-plane direction, forming the split edges (stage I). In stage I, the *g*-factor is still low because the chiral components have not developed yet. After 20 min, protruded edges (R-S boundary) split more and these tilted edges grow laterally as the overall size of the particle increases (stage II). As the chiral components of the tilted edges developed, the *g*-factor increased rapidly.

In the case of 432 helicoid II, the evolution direction of the chiral components is completely different. For the initial 30 min (stage I), the AB and A'B' edges of the rhombus ABA'B' expand with distortion. The distortion takes place gradually as a result of the increase in the R region. Distinctive edge growth was observed only after 40 min (stage II). During stage II, as the distorted edges became thicker, the chiral components were more distinguishable, increasing the *g*-factor. According to the quantification result (Extended Data Fig. 4e), the amount of adsorbed GSH also increased at this growth stage. The increasing trend of adsorbed peptides with growth time is similar to that of *g*-factor (Extended Data Fig. 4d). This finding implies that the evolution of chirality is closely related to the adsorption of GSH on the gold surface. Furthermore, different increasing trends in the *g*-factor between 432 helicoids I and II indicate that contrasting binding kinetics between Cys and GSH on the gold surface.

Numerical calculations. We analysed the optical activity of the chiral nanoparticles using a three-dimensional full-wave numerical simulation using a commercial-grade simulator (Lumerical). The calculations were based on the finite-difference time-domain (FDTD) method. The geometry of the simulation model was deduced from SEM images and a mesh was constructed non-uniformly, with a mesh size of less than 10 nm near the nanoparticles. The refractive index of water was assigned a value of 1.33 and the optical properties of gold were taken from a previous study⁴⁰.

The FDTD simulation calculates the scattering (C_{sca}) and absorption (C_{abs}) cross-sections of a given particle. The extinction cross-section ($C_{\text{ext}} = C_{\text{abs}} + C_{\text{sca}}$) is used to estimate the macroscopic absorption. According to the Beer-Lambert law, the transmission T and absorbance A through a medium of thickness l and filled with particles to a number density N is represented by $T = I/I_0 = \exp(-NlC_{\text{ext}})$ and $A = -\log_{10}(T)$. A chiral-particle medium exhibits different absorbance to left (LCP) and right (RCP) circularly polarized light (A_L and A_R); the CD calculated from this absorbance difference is approximated as

$$\text{CD} \approx (A_L - A_R) \frac{\log(10)}{4} \text{rad} = (A_L - A_R) \frac{\log(10)}{4} \frac{180}{\pi} \text{deg}$$

Here, the orientation average over 756 directions was used to account for the random orientation of the chiral nanoparticles in a water medium. The incident illumination of an electromagnetic wave travels in the $+z$ direction. Under this fixed-illumination condition, nanoparticles were rotated in three-dimensions;

the polar angle (Θ) was changed from 0° to 180° and the azimuthal angle (Φ) was simultaneously changed from 0° to 360° (Extended Data Fig. 7d). Therefore, the orientation-averaged extinction ($\langle C_{\text{ext}} \rangle_\Omega$) and CD ($\langle \text{CD} \rangle_\Omega$) could be calculated.

The electromagnetic field near the plasmonic helicoid was calculated at a normal incidence ($\Theta = 0^\circ$ and $\Phi = 0^\circ$) with a uniform mesh size of 2 nm. The electric- and magnetic-field distributions on the illuminated surface were displayed at selected wavelengths (650 nm, 950 nm and 1,200 nm); the field differences, $(|E_{\text{LCP}}|^2 - |E_{\text{RCP}}|^2)/E_0^2$ and $(|B_{\text{LCP}}|^2 - |B_{\text{RCP}}|^2)/B_0^2$, are representative of the microscopic asymmetric responses, where E_0 (B_0) indicates an amplitude of the initial electric (magnetic) field. We analysed the multipolar contribution to scattering through multipole decomposition from the calculated electromagnetic-field vectors⁴¹.

Design guidelines of chiral nanoparticles. Efforts to quantify chirality and to correlate geometric chirality with the observed chiral property have encountered many difficulties in many disciplines⁴². Despite the difficulty in correlating structural chirality with the observed macroscopic chiral properties, we tried to obtain general design guidelines for achieving large chiral responses by restricting ourselves to certain chiral particle designs (Extended Data Fig. 8). Although there are no universal design principles to intuitively predict the relationship between structural chirality and the optical chiral response, we obtained chiral responses for fixed structure designs of chiral nanoparticles using computational electrodynamics, which can be used to successfully express the optical properties of nanoparticles; that is, we can predict the morphologies of nanoparticles that will exhibit better performance. However, it is impossible to study any arbitrary design using computationally intensive full-wave numerical simulations; therefore, we show only some design guidelines for chiral nanoparticles using simplified models from SEM images (Fig. 3a, b), retaining the characteristic four-fold rotational symmetry of the helicoids.

To estimate the chiral properties of chiral nanoparticles, we first calculated the scattering cross-section and absorption cross-section of a single chiral nanoparticle at LCP and RCP incidences using FDTD with a total-field scattered-field formalism and perfectly matched layer (PML) absorbing boundaries in a water medium ($n = 1.33$). We considered random orientations of the colloidal chiral nanoparticles by rotating them to discrete orientations and averaging the results. Because the simulation of many particles with different orientations is time consuming, we simulated using a uniform 4-nm mesh. However, for some chiral nanoparticles with small feature sizes (< 20 nm), a 3-nm mesh was used. We checked some of the results against those calculated with a 1-nm mesh and found no substantial differences, although the responses obtained with a 1-nm mesh were slightly stronger.

Absorbance (abs), CD and *g*-factor are characterized as follows:

$$\begin{aligned} \text{abs} &= \frac{NI \sum C}{2 \log(10)} \\ \text{CD} &= \frac{NI \Delta C}{4} \\ g\text{-factor} &= 2 \frac{\Delta C}{\sum C} \end{aligned}$$

where $\sum C \equiv \langle C_{\text{ext,LCP}} \rangle_\Omega + \langle C_{\text{ext,RCP}} \rangle_\Omega$, $\Delta C \equiv \langle C_{\text{ext,LCP}} \rangle_\Omega - \langle C_{\text{ext,RCP}} \rangle_\Omega$; $\langle \dots \rangle_\Omega$ represents an average over all orientations, $N = 1 \times 10^{15} \text{ m}^{-3}$ is the particle number density, and $l = 1 \times 10^{-3} \text{ m}$ is the optical path length. With this definition, the *g*-factor is constrained between -2 and 2 .

It is well known that larger particles can support stronger dipole moments and higher-order modes, which can lead to stronger extinction and chiral responses. Extended Data Fig. 8a supports this claim, with chiral nanoparticles (samples 1–3) with increasing edge lengths (size of particle) in the same geometry showing increasing *g*-factors. However, to function as a metamaterial, the meta-atom, or chiral nanoparticle, should be smaller than the wavelength of the incident light; hence, the size of the chiral nanoparticle is limited to some extent. Remarkably, the optical properties of the chiral nanoparticles (both extinction and chirality) depend strongly on their subwavelength plasmonic gap. Generally, chiral nanoparticles with narrower and deeper plasmonic gaps exhibit a stronger and redshifted extinction and chiral response. The results are summarized in Extended Data Fig. 8a, in which chiral nanoparticles (samples 4–7) with increasing gap widths have decreasing *g*-factors and those (samples 8–14) with increasing gap depths have increasing *g*-factors of more than 0.7. These stronger and redshifted features may originate from a stronger dimeric coupling between the two domains separated by the plasmonic gap. This could be explained by the plasmon-hybridization model, which explains the behaviour of closely coupled plasmonic nanostructures due to the electrostatic dipole-dipole interaction resulting in an enhanced and stabilized (redshifted in wavelength) response⁴³. We also found an enhanced electric field near the plasmonic gaps in Extended Data Fig. 8b, which can result in enhanced dipole moments. This field enhancement increased as the plasmonic gap became

narrower and deeper. We also studied chiral nanoparticles (samples 16–19) with different gap angles, which is essential for the broken parity symmetry and chiral response. An achiral nanoparticle with a gap angle of 0° or 90° will not exhibit any chiral response; however, it is still difficult to quantify the structural chirality of the other chiral nanoparticles studied here.

We also simulated chiral nanoparticles with more complex geometries (Extended Data Fig. 8c). The difficulty in correlating structural chirality to the observed chiral properties is also addressed by the decreasing g -factor of chiral nanoparticles with increases in certain curvatures in samples 20–22. As stated above, structural chirality cannot be quantified and we generally rely on numerical methods or experiments to predict chiral properties. In samples 23–26, the extinction of elongated chiral nanoparticles showed noticeable changes due to an increase in size, but their g -factors remained similar despite large changes in their aspect ratio and size, of up to a factor of three. The four-fold symmetry of our chiral nanoparticle gets broken, and the orientation dependence of responses becomes larger. In sample 32, the triangular plate also showed noticeable orientation-dependent responses. This anisotropic response has commonly been observed in canonical chiral systems, such as helices, twisted-nanorods and helical arrangements of nanoparticles, and is responsible for the lowering of the average chiral response. In samples 27–31, different chiral particle designs, such as hollow chiral nanoparticles, were constructed by removing cubic domains inside the chiral nanoparticle. In samples 27 and 28, which have small void sizes, the particles did not exhibit noticeable change in their responses. However, chiral nanoparticles with large voids (samples 29–31) had g -factors of more than 0.9—the strongest g -factor of these simulations. These hollow chiral nanoparticles have very thin outer shells and exposed insides. Interestingly, strongly enhanced and redshifted responses were observed despite the greatly reduced volume.

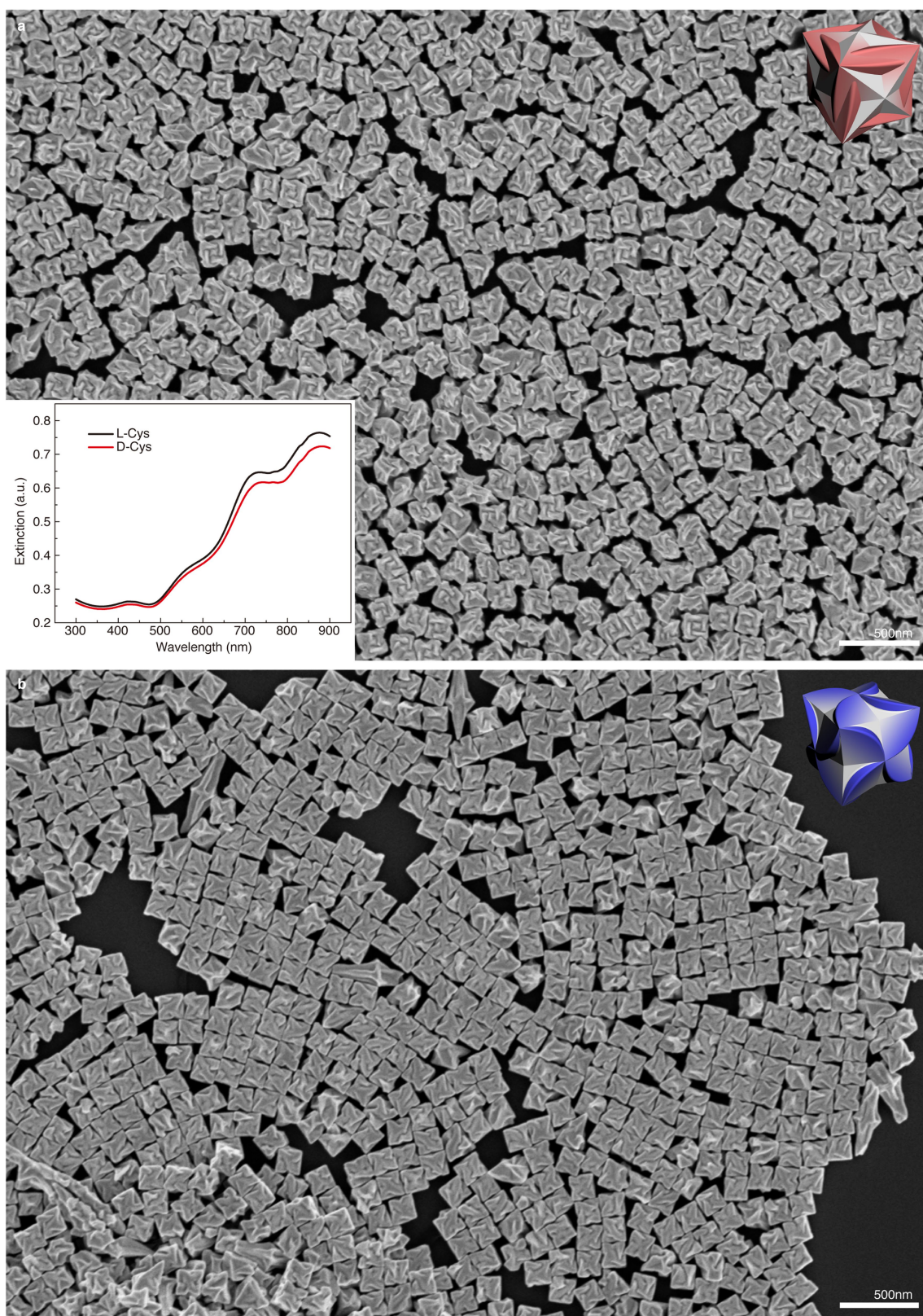
The properties of small chiral particles or chiral molecules are often explained using electric and magnetic dipole moments. This description is consistent for small chiral systems with spectrally overlapping absorption and chiral responses, which essentially limit the available g -factor. Interestingly, we observed enhanced g -factors along with enhanced CD despite increasing extinction and CD. This is because their peaks do not appear at the same wavelength and the CD peak appears between the fundamental and higher-order modes of extinction. This separation of the extinction modes appears with the redshifted response and may arise from higher-order plasmon modes. The excitation of higher-order modes using plasmon hybridization has often been reported in the field of plasmonic metamaterials⁴³ and could have a role in our chiral nanoparticles, in which nanogaps much smaller than the wavelength effectively change the overall response.

In summary, we have developed some general guidelines for designing chiral nanoparticles with high g -factors. First, both the extinction (absorbance) and chiro-optical response (CD, g -factor) depend on the size of chiral nanoparticles. This is because a larger particle supports stronger dipolar modes and even higher-order modes, resulting in stronger extinction and chiro-optical responses. Second, in general, the chiro-optical properties of chiral nanoparticles depend strongly on their 'gap'. Although the feature size of these gaps is much smaller than the wavelength, plasmonics allows considerable changes in response with subtle morphology differences. Narrower and deeper gaps allow stronger and redshifted chiro-optical responses as well as extinction, which could originate from stronger dimeric

plasmon coupling. The high-performance plasmonic chiral systems reported so far often have discrete particles that are coupled to achieve the enhanced responses. These chiral systems have linker molecules, such as DNA, to maintain their conformations. A single continuous chiral nanoparticle could have a similar property owing to the gaps that are deep and long. Therefore, designing chiral nanoparticles with high performance requires control over gap formation. Third, hollow chiral nanoparticles could achieve highly enhanced chiro-optical properties with g -factor of 0.9. In addition, the redshifted response without an increase in particle size could be beneficial because it decreases the particle-size-to-wavelength ratio, which brings this particle medium closer to the definition of a metamaterial.

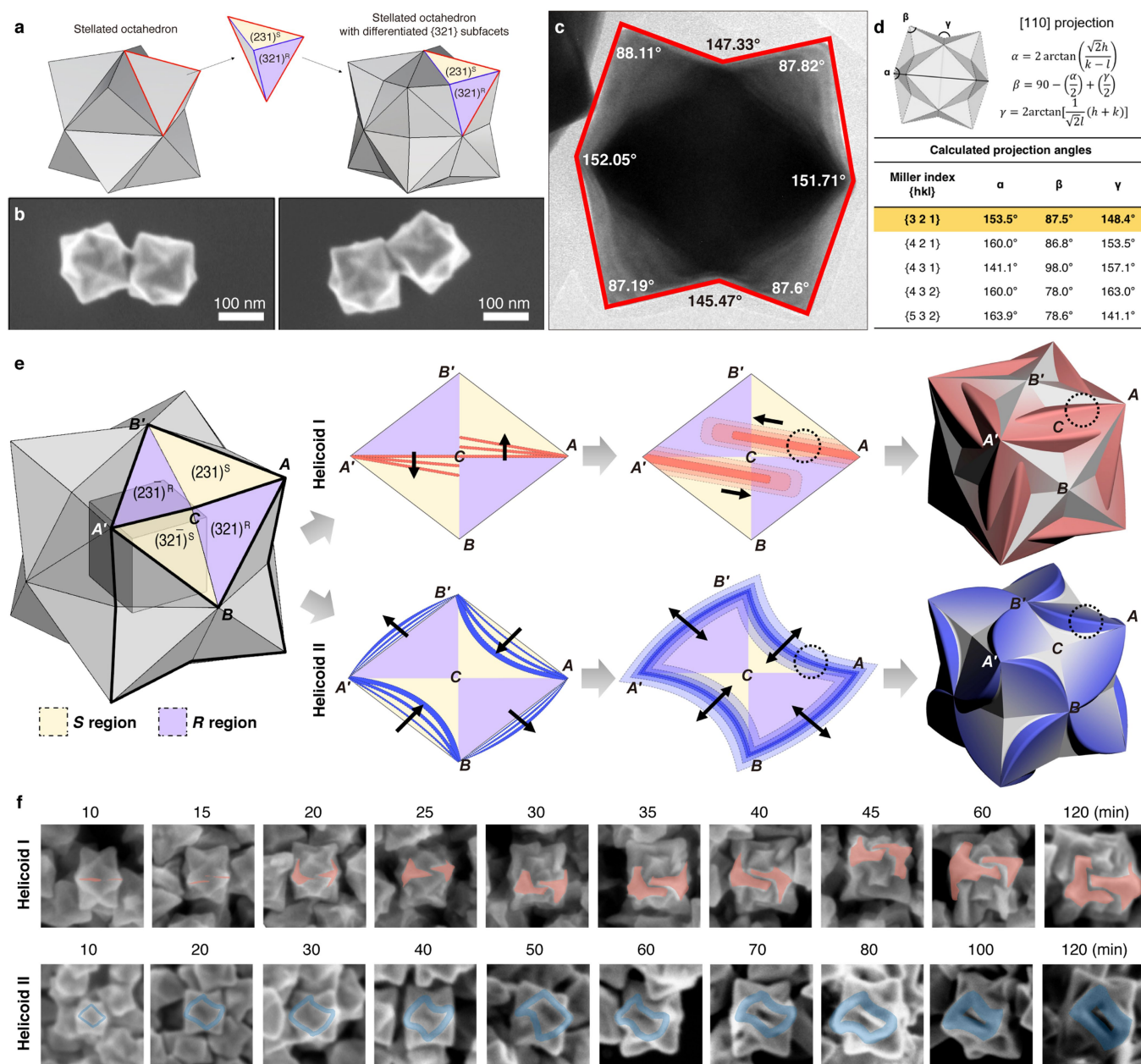
Data availability. The data that support the findings of this study are available from the corresponding authors on reasonable request.

34. Wu, H. L. et al. A comparative study of gold nanocubes, octahedra, and rhombic dodecahedra as highly sensitive SERS substrates. *Inorg. Chem.* **50**, 8106–8111 (2011).
35. Ansar, S. M. et al. Removal of molecular adsorbates on gold nanoparticles using sodium borohydride in water. *Nano Lett.* **13**, 1226–1229 (2013).
36. Yuan, M. et al. A method for removing self-assembled monolayers on gold. *Langmuir* **24**, 8707–8710 (2008).
37. Arrigan, D. W. M. & Bihan, L. L. A study of L-cysteine adsorption on gold via electrochemical desorption and copper(II) ion complexation. *Analyst* **124**, 1645–1649 (1999).
38. Bieri, M. & Burgi, T. Adsorption kinetics of L-glutathione on gold and structural changes during self-assembly: an in situ ATR-IR and QCM study. *Phys. Chem. Chem. Phys.* **8**, 513–520 (2006).
39. Bieri, M. & Bürgi, T. L-glutathione chemisorption on gold and acid/base induced structural changes: a PM-IRRAS and time-resolved in situ ATR-IR spectroscopic study. *Langmuir* **21**, 1354–1363 (2005).
40. Johnson, P. B. & Christy, R. W. Optical constants of the noble metals. *Phys. Rev. B* **6**, 4370–4379 (1972).
41. Grah, P., Shevchenko, A. & Kaivola, M. Electromagnetic multipole theory for optical nanomaterials. *New J. Phys.* **14**, 093033 (2012).
42. Barron, L. D. Chemistry: Compliments from Lord Kelvin. *Nature* **446**, 505–506 (2007).
43. Prodan, E., Radloff, C., Halas, N. J. & Nordlander, P. A hybridization model for the plasmon response of complex nanostructures. *Science* **302**, 419–422 (2003).
44. McPhie, P. Circular dichroism studies on proteins in films and in solution: estimation of secondary structure by g -factor analysis. *Anal. Biochem.* **293**, 109–119 (2001).
45. Maoz, B. M. et al. Plasmonic chiroptical response of silver nanoparticles interacting with chiral supramolecular assemblies. *J. Am. Chem. Soc.* **134**, 17807–17813 (2012).
46. Hao, C. et al. Unusual circularly polarized photocatalytic activity in nanogapped gold-silver chiroplasmonic nanostructures. *Adv. Funct. Mater.* **25**, 5816–5822 (2015).
47. Wu, X. et al. Unexpected chirality of nanoparticle dimers and ultrasensitive chiroplasmonic bioanalysis. *J. Am. Chem. Soc.* **135**, 18629–18636 (2013).
48. Yan, W. et al. Self-assembly of chiral nanoparticle pyramids with strong R/S optical activity. *J. Am. Chem. Soc.* **134**, 15114–15121 (2012).
49. Lan, X. et al. Au nanorod helical superstructures with designed chirality. *J. Am. Chem. Soc.* **137**, 457–462 (2015).
50. Yan, J., Hou, S., Ji, Y. & Wu, X. Heat-enhanced symmetry breaking in dynamic gold nanorod oligomers: the importance of interface control. *Nanoscale* **8**, 10030–10034 (2016).



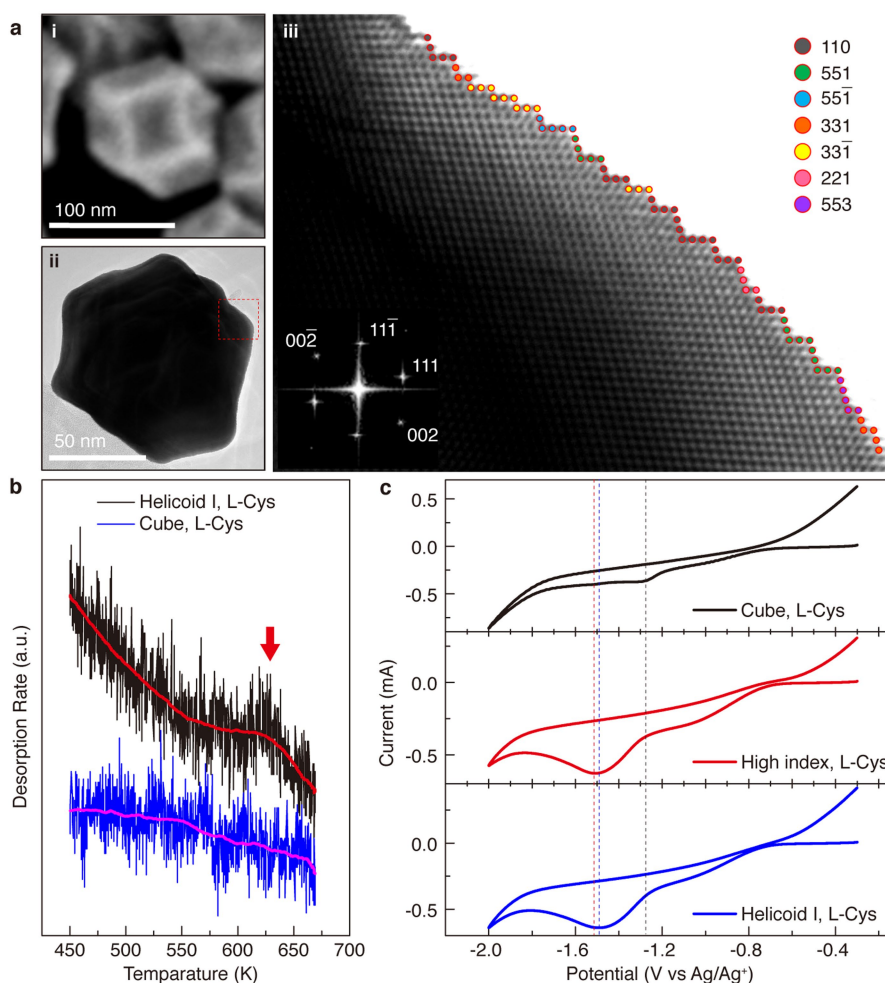
Extended Data Fig. 1 | Chiral morphology and characterization of 432 helicoids I and II. a, Large-area SEM image of 432 helicoid I. L-Cys was used as an additive. Inset, extinction spectra of 432 helicoid I synthesized

using L-Cys and D-Cys. **b,** Large-area SEM image of 432 helicoid II, synthesized with L-GSH.



Extended Data Fig. 2 | Chiral morphology development of 432 helicoids I and II. **a**, Schematic illustration of stellated octahedron with differentiated {321} facets ({321} nanoparticle). Each triangular facet of a stellated octahedron is divided into two convex {321} facets with R and S surface conformation. **b**, SEM images showing the detailed geometry of a {321} nanoparticle. **c**, Bright-field TEM image along the [110] direction showing angles (α, β, γ) between the eight outermost edges. **d**, Calculated angles between the outermost edges of an {hkl}-enclosed nanoparticle.

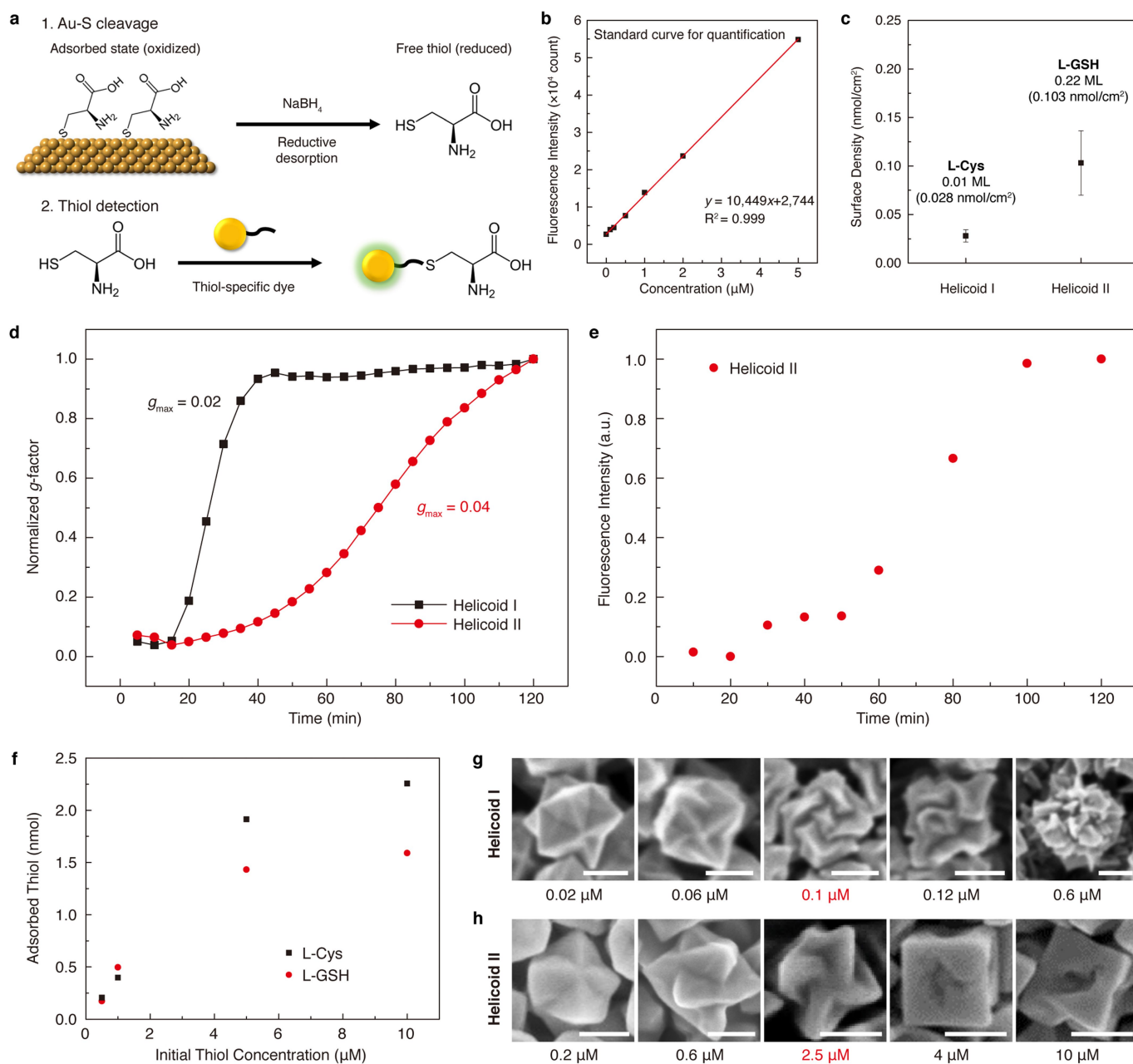
The exposed facets of the nanoparticle in **c** were indexed as {321}. **e**, Schematic illustration of the time-dependent evolution of 432 helicoids I and II. All models are viewed along the [110] direction. Starting from a {321}-indexed nanoparticle with an equal ratio of R and S regions, different R-S boundaries are split, thickened and distorted. **f**, SEM images of 432 helicoids I and II at different growth times. The chiral components that developed in 432 helicoids I and II are highlighted in red and blue, respectively.



Extended Data Fig. 3 | Interaction of L-Cys with high-index planes.

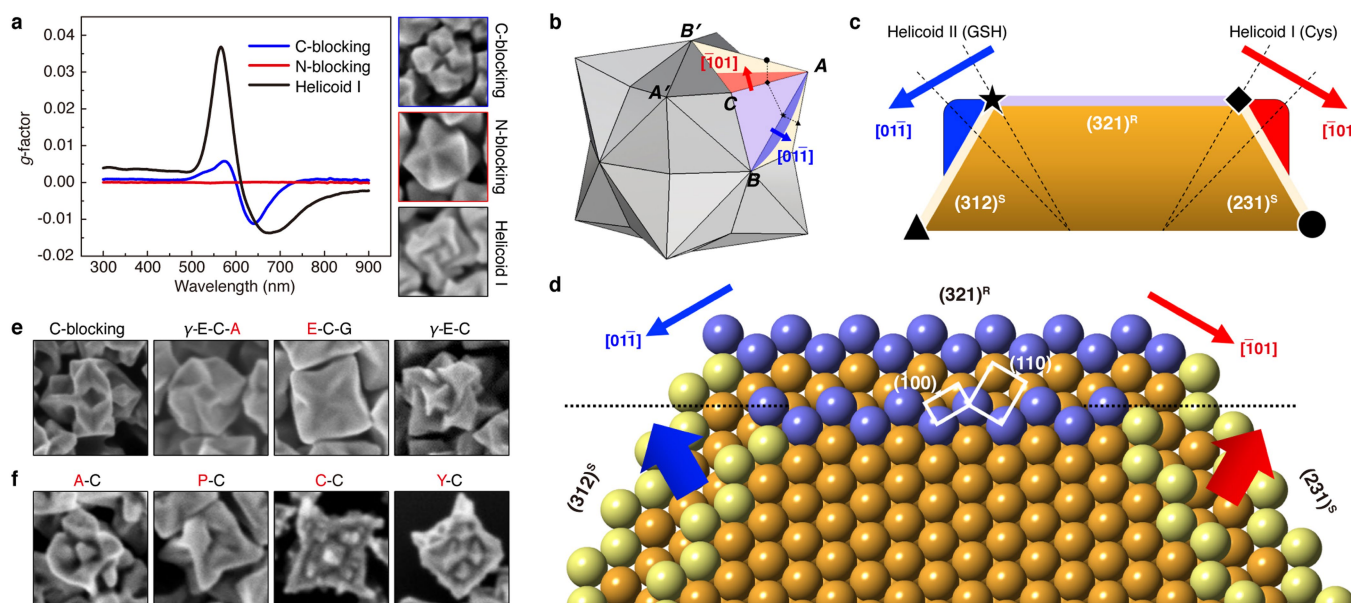
a, Atomic structure of a chiral nanoparticle at the initial stage. SEM image (i) and TEM images (ii and iii) of a chiral nanoparticle after 20 min of growth. Because the nanoparticle was oriented along the $\langle 110 \rangle$ direction, the projected boundaries in the TEM image consist of chirally distorted edges. The high-resolution TEM image of distorted edges corresponds to the red dotted box in ii. The atoms of the microfacets are marked with coloured circles, and different colours are assigned to the Miller index of each microfacet. Using microfacet nomenclature, the microstructure of (551) can be divided into three units of (111) and two units of (111̄). Inset, corresponding fast Fourier transform (FFT) showing typical patterns along the $[1\bar{1}0]$ zone. **b**, Temperature-programmed desorption spectra of L-Cys of 432 helicoid I and a low-index cubic nanoparticle, with

monitoring of CO_2 ($m/q = 44 \text{ AMU}$). As the temperature was raised at a rate of 3 K min^{-1} , helium carrier gas flowed over the dried nanoparticle sample. The distinguishable temperature-programmed desorption peak at 635 K for 432 helicoid I indicates a specific interaction of L-Cys with a kink atom on the gold surface. Cys on the cube (100) surface shows no observable peak at high temperatures. **c**, Cyclic voltammograms for a cube, a high-index stellated octahedron (with differentiated $\{321\}$ facets) and 432 helicoid I, with L-Cys measured in 0.1 M KOH-ethanol solution at a scan rate of 0.1 V s^{-1} . Negative peaks between about -1.8 V and -1.1 V originate from the reductive desorption of L-Cys by the cleavage of an Au-S bond, $\text{Au-SR} + e^- \rightarrow \text{Au} + \text{RS}^-$. Desorption peaks at more negative potentials indicate the higher adsorption energy of L-Cys on high-index gold surfaces.



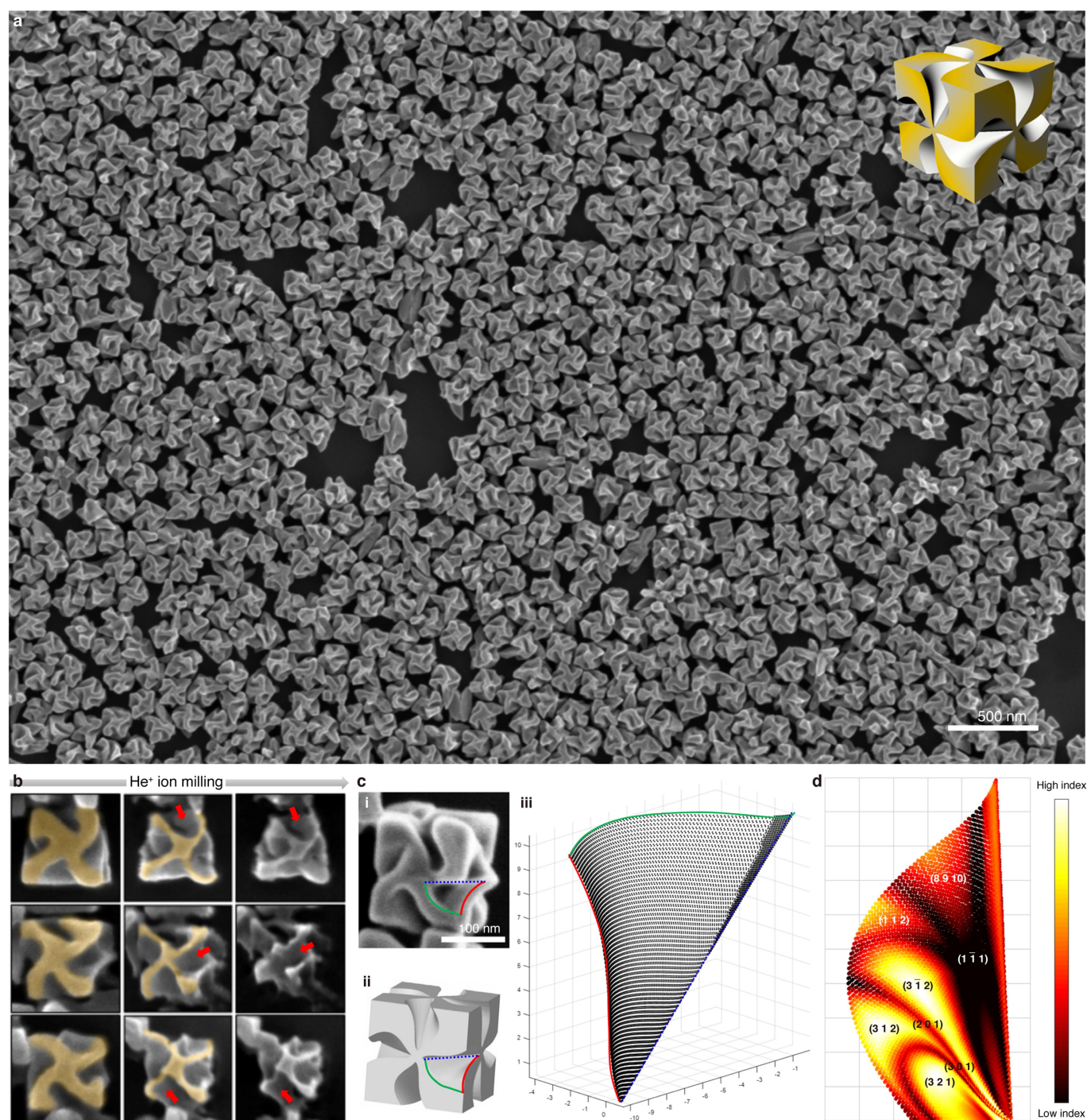
Extended Data Fig. 4 | Comparison of Cys and GSH by time-dependent concentration quantification and adsorption assay. **a**, Schematic experimental procedure for thiol quantification on the gold surface. The reduction of thiolate by NaBH_4 cleaved the Au-S bond, and the thiol group of the released molecule spontaneously reacted with the thiol-specific dye, producing a fluorescent derivative. The excitation and emission wavelengths were 405 nm and 535 nm, respectively. **b**, Concentration curve from 0 μM to 5 μM for a fluorometric assay of L-Cys. The linear fitting and corresponding R^2 value show good linearity within the measured range. **c**, Measured surface density of L-Cys and L-GSH for 432 helicoids I and II, respectively. Surface coverage is calculated using the previously reported surface densities of L-Cys and L-GSH at the fully saturated monolayer condition. Mean \pm s.d. ($n = 3$) is shown. **d**, Increase in g -factor of 432 helicoids I (Cys) and II (GSH) with time. The CD signal was measured and the normalized g -factor is displayed every 5 min during growth. The maximum g -factors (g_{max}) of 432 helicoids I and II at 120 min were

0.02 and 0.04, respectively. **e**, Amount of GSH adsorbed on 432 helicoid II at different growth times. For a detailed quantification of the amount of GSH on a nanoparticle, see Methods. **f**, Adsorption study of Cys and GSH on {321} nanoparticles. Different concentrations of Cys and GSH were added and aged for 2 h, and the amount of adsorbate was measured by subtracting the Cys and GSH concentrations in the supernatant from the initial concentration. See Methods for a detailed Cys and GSH quantification study. **g, h**, Effect of Cys and GSH concentrations on chiral morphology. SEM images of chiral nanoparticles synthesized with different concentrations of Cys (**g**) and GSH (**h**). The highest g -factor was observed at the optimum amino acid and peptide concentration (red text). At low concentrations, only achiral nanoparticles formed, but with incremental additions, chiral edges started to appear. An excess of molecule results in the overgrowth of edges and a greatly decreased CD signal, indicating that an optimal concentration exists for chirality formation. Scale bars, 100 nm.



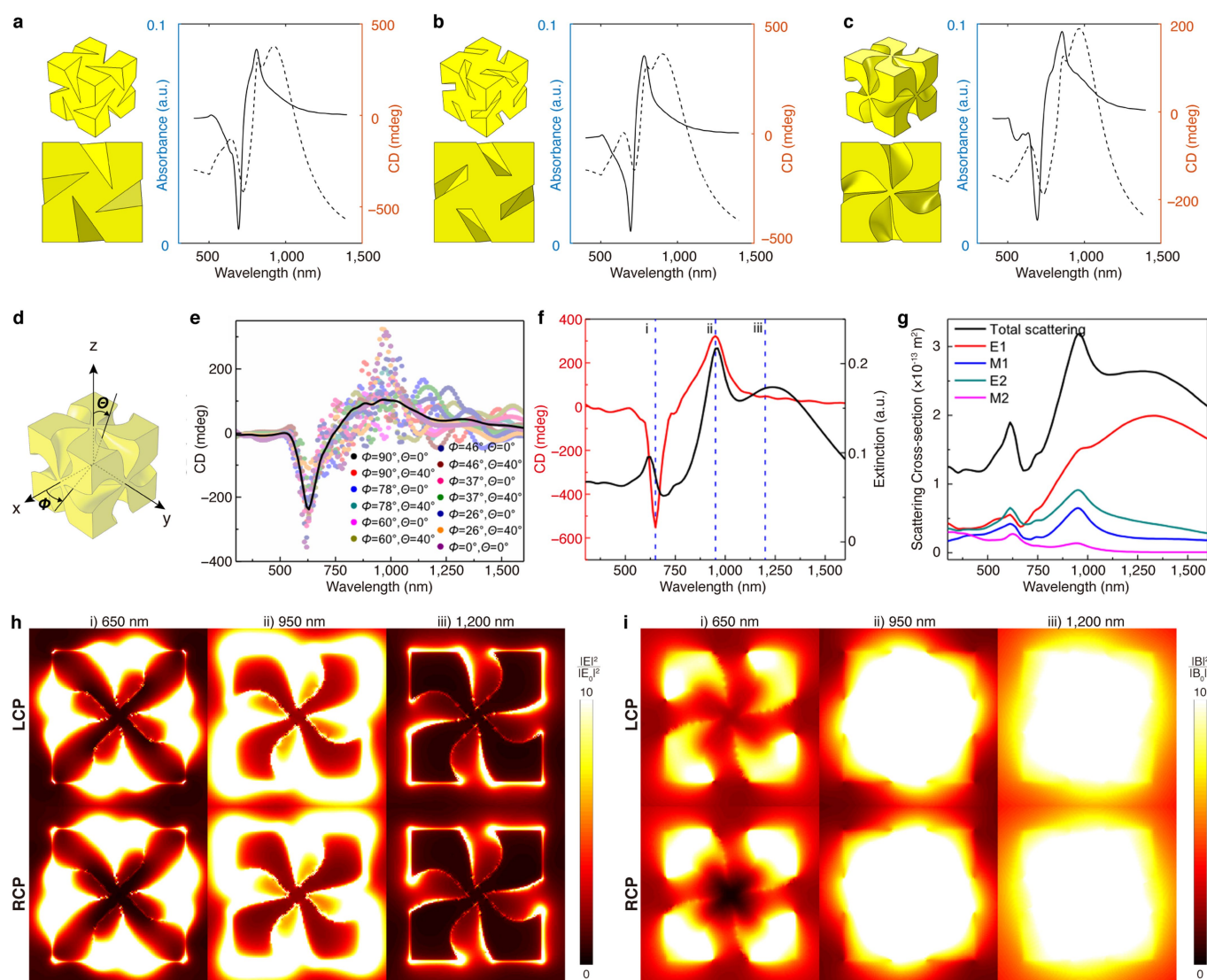
Extended Data Fig. 5 | Effect of molecular structure on chirality evolution. **a**, Effect of functional-group change in L-Cys. Comparison of g -factor and SEM images of synthesized nanoparticles with C-terminal blocked L-Cys (L-cysteine ethyl ester) (top), N-terminal blocked L-Cys (N-acetyl-L-cysteine) (middle) and L-Cys (bottom). C-blocked L-Cys changed the chiral morphology and decreased the CD intensity of the resulting nanoparticles. Furthermore, nanoparticles produced with N-blocked L-Cys showed achiral morphology without an observable CD signal. **b**, Schematic illustration of chirality formation on a {321} nanoparticle. Boundary shifts of 432 helicoids I (L-Cys) and II (L-GSH) are indicated in red and blue, respectively. **c**, Schematic (111) cross-section of $(312)^S$ – $(321)^R$ – $(231)^S$ facets. Original and newly shifted R–S boundaries are indicated with dashed lines. **d**, Atomic arrangement of

$(312)^S$ – $(321)^R$ – $(231)^S$ facets in a (111) cross-section view. The {321} surface consists of a (111) terrace and alternating {100} and {110} microfacets. The \overline{AC} boundary in 432 helicoid I shifts in the $[\overline{1}01]$ direction and the \overline{AB} boundary in 432 helicoid II shifts in the $[01\overline{1}]$ direction. The differentiated growth directions at $(312)^S$ and $(231)^S$, indicated with thick arrows, resulted in contrasting morphology for the different chiral nanoparticles. **e**, Effect of functional-group change in L-GSH. SEM images are shown of the synthesized nanoparticles prepared with L-glutathione ethyl ester (C-blocking), γ -E-C-A, E-C-G and γ -E-C sequences. **f**, SEM images of nanoparticles synthesized with different dipeptide sequences. Alanine (A), proline (P), cysteine (C) or tyrosine (Y) was added to the N terminus of L-Cys, which modified the morphology of the resulting nanoparticle substantially.



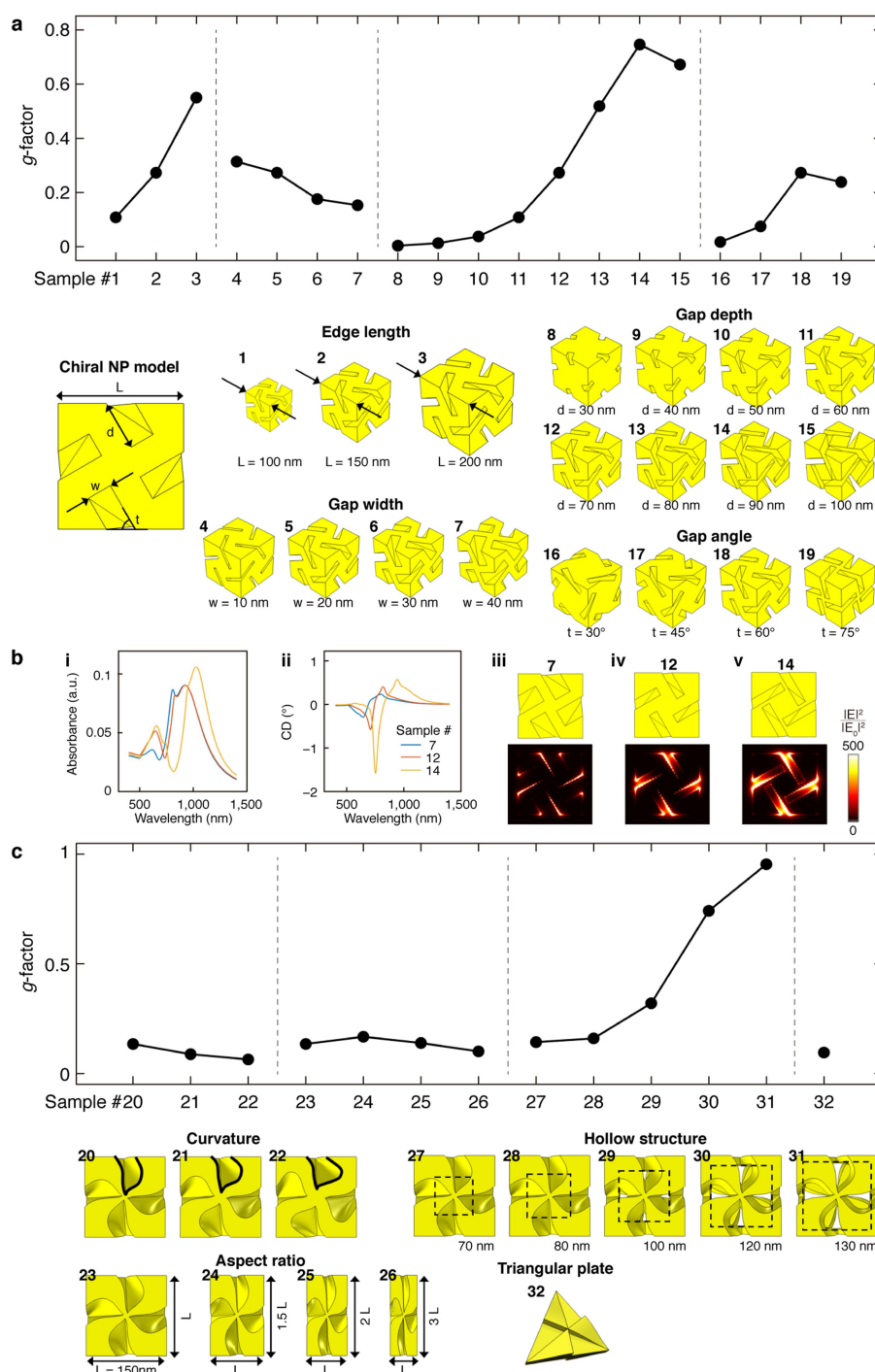
Extended Data Fig. 6 | Characterization of surfaces inside the gaps of 432 helicoid III. **a**, Large-area SEM image of 432 helicoid III nanoparticles, synthesized using an octahedral seed and L-GSH. **b**, Helium-ion microscopy secondary electron images of 432 helicoid III during the He⁺-ion milling process. The original pinwheel-like structure of 432 helicoid III is highlighted in yellow. Exposure to a He⁺-ion beam with an acceleration voltage of 30 keV and a beam current of 0.733 pA allows visualization of the interior parts of the curved surfaces, as indicated by red arrows. **c**, Modelling of the 432 helicoid III surface.

A magnified SEM image of 432 helicoid III (i), the corresponding three-dimensional model (ii) and the interpolated curved surface of 432 helicoid III (iii) are shown. The curved outlines of the chiral arm at the front and side face are indicated by green and red lines, respectively, and the internal boundary is indicated by the blue dotted lines. The three-dimensional curved-surface model of 432 helicoid III was constructed by using the interpolation of surface outlines. **d**, Distribution of Miller indices on the modelled surface. The Miller indices were calculated from a normal vector at each point on the surface.



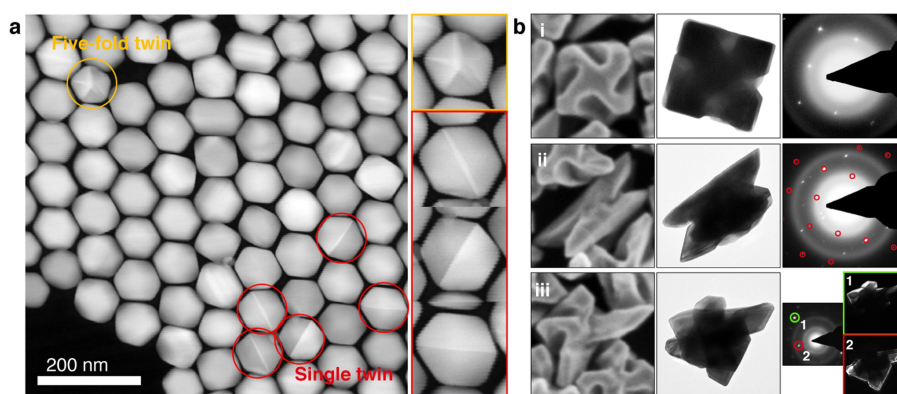
Extended Data Fig. 7 | FDTD simulation results for 432 helicoid III. **a–c**, Calculated absorbance (dashed lines) and CD (solid lines) for 432 helicoid III with triangular (**a**), rectangular (**b**) and curved (**c**) gap shapes. The corresponding three-dimensional models are displayed on the left. All three models derived from SEM images of the particles successfully reproduce the experimentally observed characteristic spectrum patterns: two main absorbance peaks, a sharp absorbance feature overlapped on the fundamental absorbance peak, a CD peak overlapped on the sharp feature, and the ‘bisignate’ CD signals, with negative peaks near 700 nm and a positive peak around 800–850 nm. This reproduction of the general features of 432 helicoid III suggests that the models can be used to study this helicoid theoretically and that they do not have to be perfect, but only need to resemble the helicoid shape sufficiently. All of the results are averaged over 756 discrete orientations and were estimated using a particle number density of $N = 10^{15} \text{ m}^{-3}$ and a cell path

length of $l = 10^{-3} \text{ m}$. **d**, Three-dimensional model and orientation of 432 helicoid III. **e**, Orientation-averaged CD spectrum ($\langle \text{CD} \rangle_{\Omega}$, black solid line) and CD spectra calculated at selected orientations (dots). $\langle \text{CD} \rangle_{\Omega}$ is averaged over 756 discrete orientations. The CD spectrum at a single orientation resembles $\langle \text{CD} \rangle_{\Omega}$ with some deviations. **f**, CD and absorbance spectra calculated with a normal incidence. **g**, Scattering cross-section decomposed by multipole analysis. The total scattering is contributed by a broad and large electric dipole mode (E1) around 1,200 nm, and a magnetic dipole (M1) and electric quadrupole (E2) around 650 nm and 950 nm near the chiro-optical peaks. A strong chiro-optical signal was observed from two other high-order modes (650 nm and 950 nm). **h, i**, Electric- and magnetic-field intensities on an illuminated helicoid surface upon normal incidence of LCP and RCP light at three different wavelengths (650 nm, 950 nm and 1,200 nm).



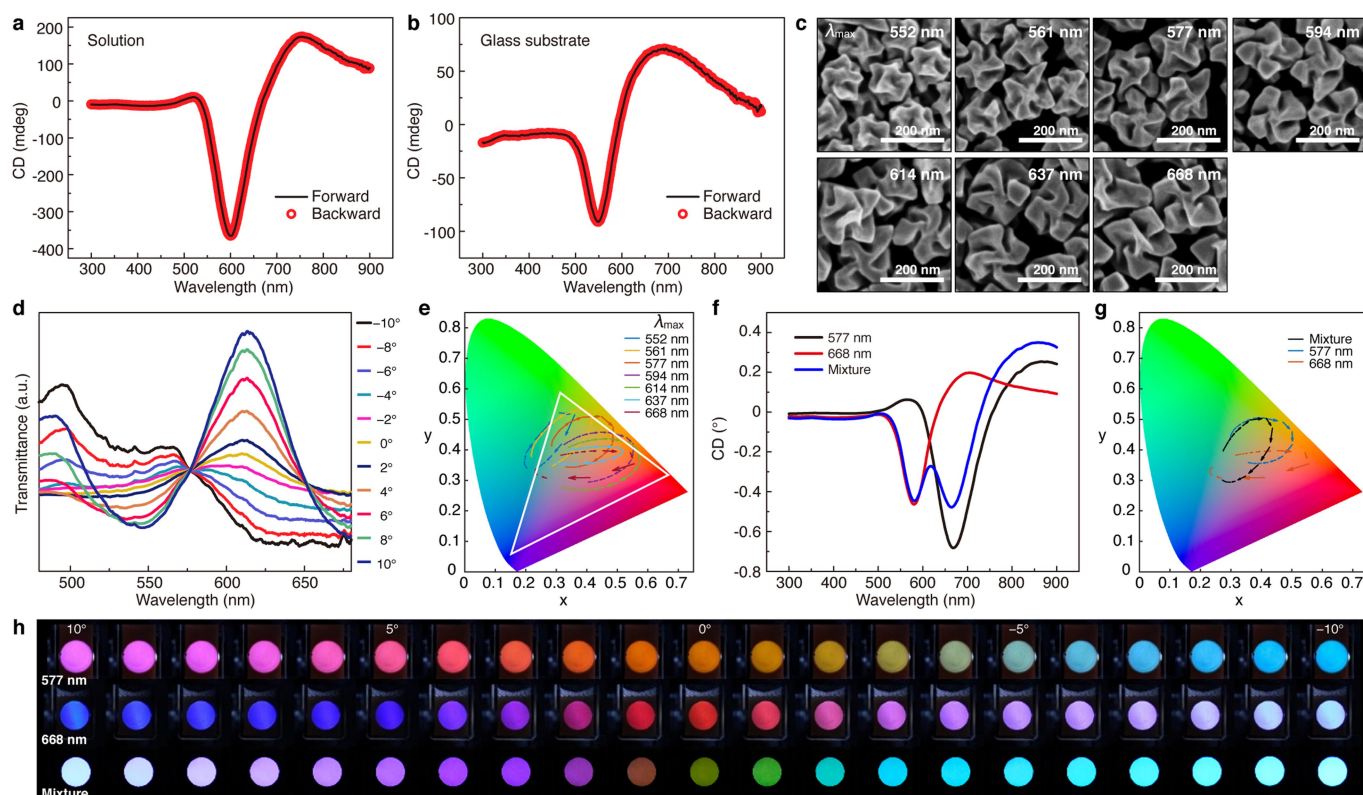
Extended Data Fig. 8 | FDTD simulations of differently modified 432 helicoid III nanoparticles to identify design guidelines. a, Calculated g -factors of chiral nanoparticles corresponding to models using parameterized chiral nanoparticles (samples 1–19). The different samples represent chiral nanoparticles with: 1–3, edge lengths L of 100–200 nm; 4–7, gap widths w of 10–40 nm; 8–15, gap depths d of 30–100 nm; or 16–19, gap angles t of 30° – 75° . The default parameters are $L = 150$ nm, $w = 20$ nm, $d = 70$ nm and $t = 60^\circ$. **b**, Calculated absorbance and CD of chiral nanoparticle samples 7 ($L = 150$ nm, $w = 40$ nm, $d = 70$ nm, $t = 60^\circ$), 12 ($L = 150$ nm, $w = 20$ nm, $d = 70$ nm, $t = 60^\circ$) and 14 ($L = 150$ nm, $w = 20$ nm, $d = 90$ nm, $t = 60^\circ$), using $N = 10^{15} \text{ m}^{-3}$ and $l = 10^{-3} \text{ m}$ (i and ii). The calculated

electric-field intensity of each of these samples on the illuminated face ($z = -75$ nm) at RCP illumination at the first CD peak—of 600 nm, 670 nm and 720 nm, respectively—is also shown (iii–v). **c**, Calculated g -factors of chiral nanoparticles corresponding to models 20–32, using chiral nanoparticles with various geometry changes: 20–22, chiral nanoparticles with increasing curvature; 23–26, chiral nanoparticles with aspect ratios of 1–3; 27–31, chiral nanoparticles with hollow structures constructed by removing cubic domains with side lengths of 70–130 nm; and 32, planar-triangle-based chiral nanoparticle with an edge length of 150 nm. The default size of chiral nanoparticles 20–31 is 150 nm.



Extended Data Fig. 9 | Effect of defects in the low-index-plane-exposed seeds on the morphology of 432 helicoid III. **a**, Characterization of twin boundary defects in seed nanoparticles. Twin boundaries were observed as bright lines in a single nanoparticle by scanning TEM imaging. Nanoparticles with a single twin and fivefold twins are indicated in red and yellow, respectively. **b**, Defect-induced morphology deformation of 432 helicoid III. SEM (left), TEM (middle) and selected-area electron diffraction (SAED; right) images are shown for an ideal 432 helicoid III (i), an irregular achiral nanoparticle (ii) and an irregular nanoparticle with broken 432 symmetry (iii). In the case of the irregular achiral particle (ii),

several diffraction spots that deviate from the regular diffraction pattern of the $\langle 100 \rangle$ zone (red) show polycrystalline character. In case of the particle with partially broken symmetry (iii), dark-field TEM images originating from diffraction spots 1 and 2 are also shown on the right, and demonstrate different crystallographic orientations in a single nanoparticle. We believe that the irregular, non-homogeneous shapes represented by ii and iii may originate from the twin boundary defects in seeds. By decreasing the population of twinned seeds, we expect that the g -factor can be further increased.



Extended Data Fig. 10 | Transmitted colour modulation by a dispersed solution of 432 helicoid III nanoparticles. **a, b,** Lorentz reciprocity of 432 helicoid III nanoparticles. The CD spectra of 432 helicoid III nanoparticles were measured from dispersion in aqueous solution (**a**) and deposition on a glass substrate (**b**). In both cases, CD measurements in the forwards and backwards directions produced identical responses. **c,** SEM images of 432 helicoid III nanoparticles with different sizes controlled by seed concentrations. Increasing the nanoparticle size resulted in a redshift in the plasmon resonance. The wavelengths at maximum CD intensity (λ_{\max}) are indicated in the images. **d,** Polarization-resolved transmittance spectra at different analyser angles. As the angle increased from -10° to 10° , transmittance at 550 nm gradually decreased, whereas that at 620 nm increased, resulting in a distinct asymmetric transition pattern. **e,** Colour

transition patterns of 432 helicoid III nanoparticles traced on CIE xy 1931 colour space (CIE, International Commission on Illumination). The white triangle indicates the RGB boundary. Each pattern shows elliptical traces with a clockwise rotational direction that reflects the asymmetric colour transition. **f,** CD spectra of a 432 helicoid III mixture. The spectral features of the broad and split CD peaks show linear superposition of the original components. **g,** Colour transition traces of the mixture on a colour space. The trace of the mixture was distinct from that of each original component and displays tailored colour transformation. **h,** Polarization-resolved transmission image of a 432 helicoid III mixture. Compared to the original components, the mixture shows different colour-transition patterns depending on the polarization angle.

Extended Data Table 1 | Comparison of *g*-factor for various chiral structures

| | Structure | Wavelength | <i>g</i> -factor | Ref. |
|--|--|------------|--------------------|-----------|
| Amino acid and peptide | L-cysteine | 215 nm | 5×10^{-3} | this work |
| | L-glutathione | 210 nm | 2×10^{-4} | this work |
| | α -helical protein | 190 nm | 3×10^{-3} | 44 |
| Chiral molecule on achiral nanoparticle | Au nanoparticle coated with peptide | 530 nm | 3×10^{-4} | 27 |
| | Ag nanoparticle coated with assembled chiral supramolecule | 530 nm | 2×10^{-3} | 45 |
| | Nanogapped Au-Ag nanoparticle | 430 nm | 1×10^{-2} | 46 |
| Chiral arrangement of multiple nanoparticles | Au-Ag nanoparticle heterodimer with antibody-antigen bridge | 400 nm | 2×10^{-2} | 47 |
| | Au nanoparticle tetrahedral superstructure with DNA-nanoparticle conjugate | 525 nm | 2×10^{-2} | 48 |
| | Au nanoparticle helical superstructure with DNA origami bundle | 700 nm | 3×10^{-2} | 9 |
| | Au nanorod helical superstructure with bifacial DNA origami sheet | 800 nm | 2×10^{-2} | 49 |
| | Twisted Au nanorod dimer with reconfigurable DNA origami bundle | 750 nm | 2×10^{-2} | 10 |
| | Twisted Au nanorod oligomer with electrostatic side-by-side assembly | 600 nm | 7×10^{-2} | 50 |
| Chiral single nanoparticle | 432 helicoid I | 565 nm | 3×10^{-2} | this work |
| | 432 helicoid II | 575 nm | 5×10^{-2} | this work |
| | 432 helicoid III | 620 nm | 2×10^{-1} | this work |

Data are from this and previous^{9,10,27,44–50} work.

High male sexual investment as a driver of extinction in fossil ostracods

Maria João Fernandes Martins^{1,5}, T. Markham Puckett², Rowan Lockwood³, John P. Swaddle⁴ & Gene Hunt^{1,5*}

Sexual selection favours traits that confer advantages in the competition for mates. In many cases, such traits are costly to produce and maintain, because the costs help to enforce the honesty of these signals and cues¹. Some evolutionary models predict that sexual selection also produces costs at the population level, which could limit the ability of populations to adapt to changing conditions and thus increase the risk of extinction^{2–4}. Other models, however, suggest that sexual selection should increase rates of adaptation and enhance the removal of deleterious mutations, thus protecting populations against extinction^{3,5,6}. Resolving the conflict between these models is not only important for explaining the history of biodiversity, but also relevant to understanding the mechanisms of the current biodiversity crisis. Previous attempts to test the conflicting predictions produced by these models have been limited to extant species and have thus relied on indirect proxies for species extinction. Here we use the informative fossil record of cytheroid ostracods—small, bivalved crustaceans with sexually dimorphic carapaces—to test how sexual selection relates to actual species extinction. We show that species with more pronounced sexual dimorphism, indicating the highest levels of male investment in reproduction, had estimated extinction rates that were ten times higher than those of the species with the lowest investment. These results indicate that sexual selection can be a substantial risk factor for extinction.

Sexual selection favours traits that confer advantages to competition for access to mates, often leading to the evolution of costly, exaggerated characteristics^{7,8}. The evolutionary costs of such traits help to enforce the honesty of the associated displays¹, but can also reduce fitness of populations in general and thereby increase the risk of population extinction in response to environmental change^{2–4}. Alternatively, sexual selection could instead reinforce natural selection, more effectively remove deleterious mutations, and thereby speed up adaptation, which could decrease the risk of extinction^{3,5,6,9}. The conflicting predictions generated by these two types of evolutionary models have prompted empirical tests of the relationship between sexual selection and extinction risk. Experiments on laboratory populations have found adaptation to be more effective in the presence of sexual selection in some cases^{5,10}, but not others¹¹. Studies of wild populations have not found evidence that sexual selection protects against extinction, but have instead suggested that it either increases extinction risk^{12–15} or that it has no effect^{16–18}. Notably, all of these studies have examined extant species and have therefore been limited to studying indirect proxies of extinction rather than true lineage terminations. Such proxies have included population decline^{17,18}, local extirpation^{12,14,19} and conservation status^{13,16}. Because the models predict evolutionary outcomes, it is important that we investigate patterns of actual species extinction in association with changes in the strength of sexual selection.

The fossil record has documented the origin, persistence and extinction of a large number of species. Palaeontologists routinely compare the longevity of fossil taxa to test factors that have been hypothesized to increase or decrease extinction risk^{20,21}. These approaches have not

yet been applied to sexual selection, because males and females can seldom be distinguished in fossil remains and, therefore, we usually know very little about sexual dimorphism and sexual selection in extinct species²². Cytheroid ostracods, however, are a notable exception to this rule. Males in extant members of this superfamily can be distinguished from females by their relatively elongated carapaces²³ (Fig. 1). This shape difference arises from an expansion of the posterior region that accommodates the large sperm pumping and copulatory apparatus of males²³. Because this difference is expressed in the mineralized and readily preserved carapace, sexes can be discerned even in extinct populations. Reports from living cytheroids have suggested that sexual differences in carapace size and shape can reflect differences in male investment in reproduction: males with larger carapaces bear

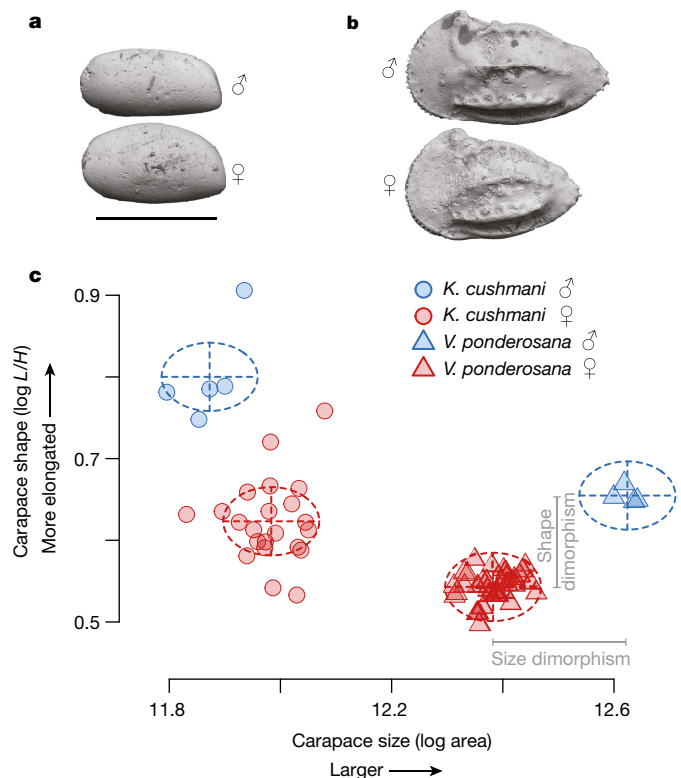


Fig. 1 | Sexual dimorphism in two species of cytheroid ostracods. **a, b**, Example males (top) and females (bottom) of *Kriehleia cushmani* (**a**) and *Veenia ponderosana* (**b**). **c**, Carapace size versus shape (circles, *K. cushmani*, $n = 27$; triangles, *V. ponderosana*, $n = 39$) with separate sex clusters for each species (blue, males, red, females). Magnitudes of sexual dimorphism were computed as male minus female means. Scale bar, 200 μm (applies to all specimens). These fossils were sampled from the Marlbrook Marl (*K. cushmani*) and Annona Chalk (*V. ponderosana*) in Arkansas, USA.

¹Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. ²Department of Geography and Geology, The University of Southern Mississippi, Hattiesburg, MS, USA. ³Department of Geology, The College of William and Mary, Williamsburg, VA, USA. ⁴Department of Biology, The College of William and Mary, Williamsburg, VA, USA. ⁵These authors contributed equally: Maria João Fernandes Martins, Gene Hunt. *e-mail: hunte@si.edu

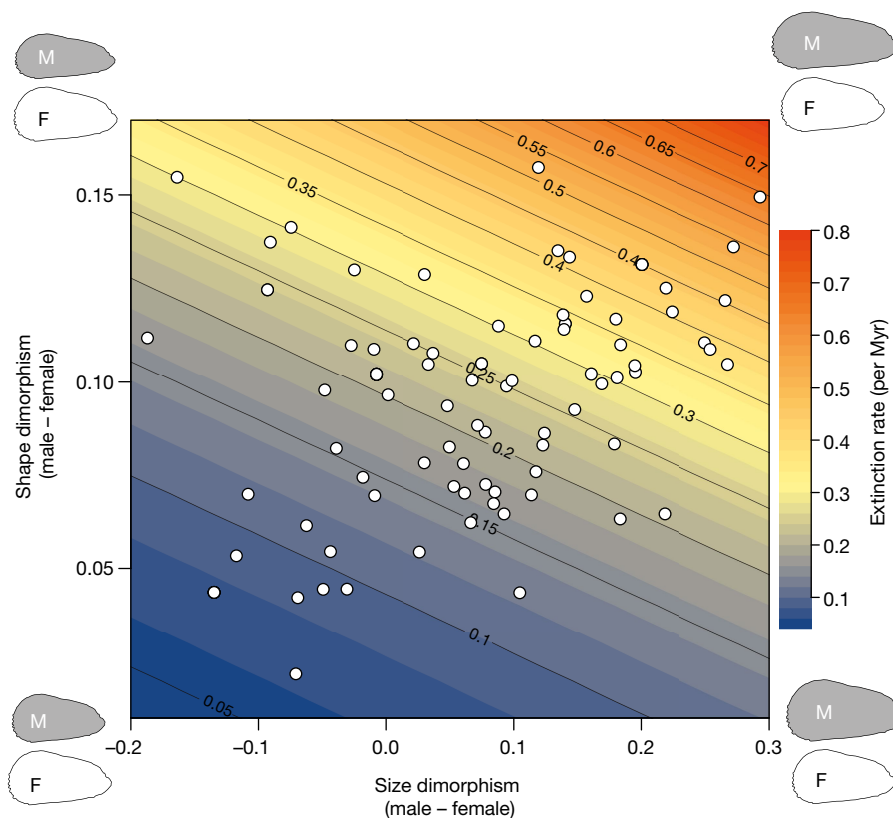


Fig. 2 | Model-predicted extinction rate according to sexual size and shape dimorphism. Magnitudes of sexual size and shape dimorphism were computed as male minus female means. Each dot represents the size and shape dimorphism of a species ($n = 93$), with colour contours

corresponding to per-Myr extinction rates predicted by capture-mark-recapture modelling. Silhouettes illustrate sexual dimorphism patterns at each corner of the plot (male, grey; female, white).

disproportionately larger sex organs²⁴ and the relative elongation of males can be related to the relative size of their copulatory organs²⁵.

Recent work has comprehensively documented magnitudes of sexual size and shape dimorphism in cytheroid ostracod fauna from the Late Cretaceous epoch (approximately 84–66 million years ago) of the US coastal plain²⁶. Species of this fauna vary greatly in their sexual dimorphism: males range from 30% larger to 20% smaller than females, with abundant variation in shape dimorphism as well (Fig. 2). To test whether this large variation in male investment among species has consequences for extinction risk, we combined these data with a high-resolution study of the stratigraphic occurrences of 93 species in Late Cretaceous strata in eastern Mississippi²⁷ (Extended Data Fig. 1). Using capture-mark-recapture methods²⁸, we fitted a series of 576 models in which probabilities of extinction, speciation and preservation are constant over time, variable over time or dependent on covariates, such as the magnitude of sexual dimorphism or on other traits that may be related to evolutionary outcomes. The key assessment of the influence of sexual selection on extinction hinges on comparisons between models in which extinction depends on sexual dimorphism versus those in which it does not.

The fits of these models strongly indicate that extinction probabilities increase with male reproductive investment as reflected by sexual dimorphism. Only twenty models receive non-trivial support²⁹ (difference in corrected Akaike information criterion (ΔAICc) < 10 ; higher values indicate lower support), and all of these models except for one have extinction probabilities that depend on sexual dimorphism in size, shape or both (Table 1; full model results are in Supplementary Table 1). Support for the best model in which extinction is independent of sexual dimorphism is almost negligible compared to that of the best-supported model (model 18, $\Delta\text{AICc} = 9.25$). Overall, models in which extinction depends on sexual dimorphism collectively account for 99.3% of the available model support (that is, Akaike weight).

Estimated coefficients averaged across models indicate that extinction risk increases markedly with size and shape dimorphism (Fig. 2): predicted extinction rates are approximately tenfold higher for the most dimorphic species (0.64 per million years (Myr^{-1})) compared to the species with dimorphism that are indicative of the lowest levels of male investment in reproduction (0.06 Myr^{-1}). These differences in extinction rate correspond to expected species durations of 1.6 and 15.5 Myr, respectively. The similarity of estimated coefficients across models (Extended Data Fig. 2) emphasizes the consistent signal of increased extinction risk in taxa with males that are larger and more elongated than females (see also Extended Data Fig. 3).

Extinction in the best-supported model increases with shape dimorphism (Table 1), and the model-averaged 95% confidence interval for this coefficient excludes zero (Extended Data Fig. 2). Size dimorphism does affect extinction in some well-supported models (Table 1), but the effect is less consistent (Extended Data Fig. 2) and these data cannot decisively determine whether extinction risk increases only with size dimorphism, only with shape dimorphism or with both. Previous work³⁰ has suggested that speciation might also be facilitated by sexual selection, but we find little evidence for this relationship: speciation probabilities increase with magnitudes of sexual dimorphism in some models, but not in any of the ones with the highest support (Table 1).

Behavioural observations from living cytheroids show no indication that the sexual dimorphism of the carapace is related to pre-copulatory signalling to females or to direct contests among males³¹. Rather, it is more likely that this dimorphism reflects investment in sexual reproduction itself. In extant species of the cytheroid genus *Cyprideis*, sexual size dimorphism is correlated with the size of the male genitalia. The strongest correlations with size involve the large, muscular sperm pump²⁴, suggesting that size dimorphism might relate to the quantity, size or transfer efficiency of sperm. Resources devoted to sperm competition are unavailable for other functions needed for survival

Table 1 | Best-supported models for speciation and extinction

| Rank | Extinction | Speciation | ΔAICc |
|------|---|--|---------------------|
| 1 | DM _{shape} + occupancy + family | Occupancy + family | 0.00 |
| 2 | DM _{size} + DM _{shape} + occupancy + family | Occupancy + family | 1.00 |
| 3 | DM _{shape} + occupancy + family | Occupancy | 2.63 |
| 4 | DM _{size} + DM _{shape} + occupancy | Occupancy + family | 3.32 |
| 5 | DM _{size} + DM _{shape} + occupancy + family | Occupancy | 3.38 |
| 6 | DM _{shape} + occupancy + family | Constant | 4.82 |
| 7 | DM _{size} + occupancy + family | Occupancy + family | 4.85 |
| 8 | DM _{size} + DM _{shape} + occupancy + family | Constant | 6.06 |
| 9 | DM _{size} + occupancy + family | Occupancy | 6.56 |
| 10 | DM _{shape} + occupancy + family | DM _{size} | 6.66 |
| 11 | DM _{shape} + occupancy + family | DM _{shape} | 6.86 |
| 12 | DM _{shape} + occupancy + family | Family | 7.09 |
| 13 | DM _{size} + DM _{shape} + occupancy + family | DM _{size} | 7.47 |
| 14 | DM _{size} + DM _{shape} + occupancy | Occupancy | 7.80 |
| 15 | DM _{size} + DM _{shape} + occupancy + family | DM _{shape} | 8.11 |
| 16 | DM _{size} + DM _{shape} + occupancy + family | Family | 8.46 |
| 17 | DM _{shape} + occupancy + family | DM _{size} + DM _{shape} | 8.73 |
| 18 | Occupancy + family | Occupancy + family | 9.25 |
| 19 | DM _{size} + DM _{shape} + occupancy + family | DM _{size} + DM _{shape} | 9.53 |
| 20 | DM _{size} + occupancy + family | Constant | 9.80 |

Models are listed in order of decreasing model support as measured by ΔAICc . Under extinction and speciation the covariates are listed that modify speciation and extinction probabilities for those models. The top twenty models include all of those with non-negligible²⁹ model support ($\Delta\text{AICc} < 10$); in all of these models except for one (model 18), extinction risk depends on sexual dimorphism. DM_{size}, sexual size dimorphism; DM_{shape}, sexual shape dimorphism. Occupancy measures how widely a species is found and family indicates the taxonomic family. Constant indicates that speciation or extinction probabilities are the same in each time interval and do not depend on sexual dimorphism or any other covariate. Results for the full set of models are given in Supplementary Table 1.

and ejaculates themselves may be costly to produce³². Increased sperm competition may also be harmful to females who, in turn, may evolve increasingly costly counter-adaptations³³. Therefore, the dimorphism that we have documented here is likely costly at the population level and could contribute to the increased extinction risk in high-investment species. Sexual selection may also indirectly increase extinction risk by pulling male and female phenotypes away from their natural selection optima² or by lowering the effective population size through skewed reproductive success³.

Palaeontologists have documented a variety of factors that can contribute to lineage extinction²¹. The most consistent finding in these studies is that widespread and abundant taxa tend to have a lower extinction risk²¹, and indeed, we also find that our proxy for these characteristics, occupancy, is an important predictor of extinction risk, with high occupancy protecting against extinction (Table 1). Other palaeontological studies have reported that extinction and origination rates can vary markedly across taxa³⁴ and we also find that these rates differ across taxonomic families (Table 1). The capture–mark–recapture approach accounts for these substantial contributions to extinction risk, but shows that these factors on their own cannot explain the data as well as models that also include sexual dimorphism (Table 1).

We have assessed still other potential predictors of extinction risk, but none of these predictors can account for the relationship between extinction and sexual dimorphism that we document here. Carapace size and shape are only weakly related to sexual dimorphism²⁶, and substituting these factors for dimorphism in the best supported model greatly reduces support ($\Delta\text{AICc} = 9.64$). Stratigraphic architecture can have a strong effect on the distribution of observed extinctions³⁵, but there is no reason to expect it to differently affect species according to their sexual dimorphism. Moreover, we have repeated the analyses here using occurrence data from a different composite reference section several hundred kilometres away in central Alabama (Extended Data Fig. 1) and obtained similar results (Extended Data Table 1).

Current extinction risks are heavily shaped by human impacts and their drivers may differ from extinctions in pre-human ecosystems³⁶. Nevertheless, if costly male traits increase extinction risk by decreasing the capacity of populations to respond to changing conditions, this mechanism should also operate in present-day populations and thus compound risks from habitat destruction, invasive species, climate

change and other anthropogenic causes. Moreover, if the effect of sexual dimorphism on extinction is as strong in other taxa as what we document here for cytheroid ostracods, intense sexual selection may be important for attempts to evaluate extinction risk and design management plans of extant species.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0020-7>.

Received: 6 November 2017; Accepted: 16 February 2018.

Published online 11 April 2018.

- Höglund, J. & Sheldon, B. C. The cost of reproduction and sexual selection. *Oikos* **83**, 478–483 (1998).
- Lande, R. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* **34**, 292–305 (1980).
- Kokko, H. & Brooks, R. Sexy to die for? Sexual selection and the risk of extinction. *Ann. Zool. Fenn.* **40**, 207–219 (2003).
- Tanaka, Y. Sexual selection enhances population extinction in a changing environment. *J. Theor. Biol.* **180**, 197–206 (1996).
- Lumley, A. J. et al. Sexual selection protects against extinction. *Nature* **522**, 470–473 (2015).
- Lorch, P. D., Proulx, S., Rowe, L. & Day, T. Condition-dependent sexual selection can accelerate adaptation. *Evol. Ecol. Res.* **5**, 867–881 (2003).
- Darwin, C. *The Descent of Man, and Selection in Relation to Sex*. (John Murray, London, 1871).
- Andersson, M. *Sexual Selection*. (Princeton Univ. Press, Princeton, 1994).
- Martínez-Ruiz, C. & Knell, R. J. Sexual selection can both increase and decrease extinction probability: reconciling demographic and evolutionary factors. *J. Anim. Ecol.* **86**, 117–127 (2017).
- Jacomb, F., Marsh, J. & Holman, L. Sexual selection expedites the evolution of pesticide resistance. *Evolution* **70**, 2746–2751 (2016).
- Reding, L. P., Swaddle, J. P. & Murphy, H. A. Sexual selection hinders adaptation in experimental populations of yeast. *Biol. Lett.* **9**, 20121202 (2013).
- Doherty, P. F. Jr et al. Sexual selection affects local extinction and turnover in bird communities. *Proc. Natl Acad. Sci. USA* **100**, 5858–5862 (2003).
- Bro-Jørgensen, J. Will their armaments be their downfall? Large horn size increases extinction risk in bovids. *Anim. Conserv.* **17**, 80–87 (2014).
- Sorci, G., Møller, A. P. & Clobert, J. Plumage dichromatism of birds predicts introduction success in New Zealand. *J. Anim. Ecol.* **67**, 263–269 (1998).
- McLain, D. K., Moulton, M. P. & Sanderson, J. G. Sexual selection and extinction: the fate of plumage-dimorphic and plumage-monomorphic birds introduced onto islands. *Evol. Ecol. Res.* **1**, 549–565 (1999).
- Morrow, E. H. & Fricke, C. Sexual selection and the risk of extinction in mammals. *Proc. R. Soc. B* **271**, 2395–2401 (2004).
- Prinzing, A., Brändle, M., Pfeifer, R. & Brandl, R. Does sexual selection influence population trends in European birds? *Evol. Ecol. Res.* **4**, 49–60 (2002).
- Thomas, G. H., Lanctot, R. B. & Székely, T. Can intrinsic factors explain population declines in North American breeding shorebirds? A comparative analysis. *Anim. Conserv.* **9**, 252–258 (2006).
- McLain, D. K., Moulton, M. P. & Redfearn, T. P. Sexual selection and the risk of extinction of introduced birds on oceanic islands. *Oikos* **74**, 27–34 (1995).
- Payne, J. L. et al. Extinction intensity, selectivity and their combined macroevolutionary influence in the fossil record. *Biol. Lett.* **12**, 20160202 (2016).
- Orzechowski, E. A. et al. Marine extinction risk shaped by trait–environment interactions over 500 million years. *Glob. Change Biol.* **21**, 3595–3607 (2015).
- Knell, R. J., Naish, D., Tomkins, J. L. & Hone, D. W. E. Sexual selection in prehistoric animals: detection and implications. *Trends Ecol. Evol.* **28**, 38–47 (2013).
- Cohen, A. C. & Morin, J. G. Patterns of reproduction in ostracodes: a review. *J. Crustac. Biol.* **10**, 184–212 (1990).
- Martins, M. J. F., Hunt, G., Lockwood, R., Swaddle, J. P. & Horne, D. J. Correlation between investment in sexual traits and valve sexual dimorphism in Cyprideis species (Ostracoda). *PLoS ONE* **12**, e0177791 (2017).
- Kamiya, T. Heterochronic dimorphism of *Loxocoelha uranouchiensis* (Ostracoda) and its implications for speciation. *Paleobiology* **18**, 221–236 (1992).
- Hunt, G. et al. Sexual dimorphism and sexual selection in cytheroidean ostracodes from the Late Cretaceous of the U.S. coastal plain. *Paleobiology* **43**, 620–641 (2017).
- Puckett, T. M. Santonian–Maastrichtian planktonic foraminiferal and ostracode biostratigraphy of the northern Gulf Coastal Plain, USA. *Stratigraphy* **2**, 117–146 (2005).
- Liow, L. H. & Nichols, J. D. in *Quantitative Methods in Paleobiology* The Paleontological Society Papers Vol. 16 (eds Alroy, J. & Hunt, G.) 81–94 (The Paleontological Society, New Haven, 2010).
- Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference*. 2nd edn (Springer, New York, 2010).
- Servedio, M. R. & Boughman, J. W. The role of sexual selection in local adaptation and speciation. *Annu. Rev. Ecol. Syst.* **48**, 85–109 (2017).

31. Horne, D. J., Danielopol, D. L. & Martens, K. in *Sex and Parthenogenesis. Evolutionary Ecology of Reproductive Modes in Non-marine Ostracods* (ed. Martens, K.) 157–195 (Backhuys Publishers, Leiden, 1998).
32. Lüpold, S. et al. How sexual selection can drive the evolution of costly sperm ornamentation. *Nature* **533**, 535–538 (2016).
33. Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373**, 241–244 (1995).
34. Stanley, S. M. An analysis of the history of marine animal diversity. *Paleobiology* **33**, 1–55 (2007).
35. Holland, S. M. The stratigraphic distribution of fossils. *Paleobiology* **21**, 92–109 (1995).
36. Payne, J. L., Bush, A. M., Heim, N. A., Knope, M. L. & McCauley, D. J. Ecological selectivity of the emerging mass extinction in the oceans. *Science* **353**, 1284–1286 (2016).

Acknowledgements We thank L. Smith (LSU) and C. Sanford (NMNH) for help with access to museum specimens and C. Hall, C. Sweeney, J. Shaw, and J. Stedman for assistance in data collection. This research was supported by NSF-EAR 1424906 and the National Museum of Natural History, Smithsonian Institution.

Reviewer information *Nature* thanks M. Gage, R. Knell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions G.H. and M.J.F.M. collected the sexual dimorphism data and performed the analyses; T.M.P. collected the stratigraphic data and helped with the collection of dimorphism data. R.L., J.P.S. and G.H. designed the project and M.J.F.M. and G.H. primarily wrote the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0020-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0020-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to G.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Dimorphism data. Procedures for measuring sexual dimorphism in valve size and shape have been described previously²⁶. In brief, we photographed individual ostracods from field and museum collections and computed body size (area) and shape (length-to-height ratio, L/H) from their digitized outlines. Sex clusters were recognized from the log-transformed size and shape data using finite mixture models, with the more elongated (higher L/H) cluster interpreted to be male similar to living cytheroids (Fig. 1). We computed magnitudes of size and shape dimorphism as male minus female means in $\log(\text{area})$ and $\log(L/H)$, respectively. Because males are the more elongated sex in cytheroids, shape dimorphism is always positive. By contrast, size dimorphism can vary in magnitude and direction: it is positive when males are larger than females and negative when females are the larger sex. **Stratigraphic data.** We analysed stratigraphic occurrence data for 93 ostracod species from a composite reference section in eastern Mississippi²⁷ (Extended Data Fig. 1), omitting non-cytheroid ostracods and taxa that have not been resolved to species level. We recorded the presence or absence of these species for 88 samples for which the stratigraphic position relative to several marker beds was measured. We combined some adjacent samples with low abundances to yield 71 samples for analysis and converted stratigraphic heights to absolute ages using an age model with tiepoints from the range endpoints of several planktonic foraminifera (*Dicarinalla asymetrica* (last appearance) 83.64 million years ago (Ma), *Radotruncana calcarata* (first appearance) 76.18 Ma, *Radotruncana calcarata* (last appearance) 75.71 Ma, *Globotruncana aegyptiaca* (first appearance) 74 Ma) and setting the youngest Cretaceous sediments in the section to be 66.3 Ma, following a previously published study³⁷. The resulting 71 samples spanned over 200 m of section and 17.5 Myr with a mean spacing between consecutive samples of 250 thousand years (median = 122 thousand years) (Extended Data Fig. 1).

Of the 93 species analysed, we had direct estimates of sexual dimorphism for 69. The remaining species were found too rarely to infer sex clusters and were assigned dimorphism values equal to the mean of their congeneric species (17 species); if no data were available from a genus, we used family means instead (7 species). These substitutions are reasonable, because there is phylogenetic signal in sexual dimorphism in this fauna²⁶.

We also analysed stratigraphic occurrence data from an additional composite reference section in central Alabama (Extended Data Fig. 1) as a replicate to assess the robustness of the results from the eastern Mississippi section (Extended Data Table 1).

Modelling extinction, speciation and preservation. In order to model speciation, extinction and preservation probabilities of species, we used capture–mark–recapture (CMR) methods implemented in the program MARK³⁸ using the interface provided by the R package RMark³⁹. The input data for CMR are the set of encounter histories for all species. Each encounter history is represented by a vector with an entry for each sample, with ‘1’ indicating that the focal species was sampled and ‘0’ indicating that it was not (for example, 001101000 for a species that was absent in the first two samples, present in the third, fourth and sixth samples, and then not encountered thereafter). Such encounter histories allow calculation of probabilities of preservation, origination and extinction using the Pradel seniority model (following a previously published study⁴⁰). MARK uses maximum likelihood to estimate origination and extinction from first and last occurrences in encounter histories while allowing for incomplete sampling.

The CMR approach has several strengths for the present study compared to alternative approaches. First is the fact that one does not need to observe the entire temporal range of species to fit the models. Some species existed before our window of observation and others persisted after, but this is handled without issue in the CMR framework. Second, CMR approaches estimate speciation and extinction probabilities while accounting for incomplete and potentially variable preservation. Some alternatives, such as survival analysis or the analysis of raw stratigraphic ranges do not have these benefits.

The third advantage of CMR is perhaps the most important for the present study: it allows for parameters to be functions of covariates, which can be attributes of either samples or taxa. We modelled speciation and extinction probabilities as functions of size dimorphism, shape dimorphism or both. As alternatives, we also considered models in which these probabilities were constant (the same for all species and time intervals) and variable over time (estimated separately for each time interval). Preservation probabilities were similarly modelled as constant, variable with time, variable across each geological formation and member, and as a function of log-sample size.

In addition, we considered two additional variables that are known to influence speciation and extinction in other palaeontological studies: occupancy and family

membership. Occupancy is a common measurement of how widespread a taxon is⁴¹, here calculated as the proportion of samples in which a species was found to occur, excluding samples from formations for which a taxon has never been found in order to omit samples from before it originated or after it went extinct. We also excluded samples from the focal composite reference section so that the occupancy data would be independent of the observations used for CMR modelling. Taxonomic family was considered to capture variation in speciation and extinction across broader clades; phylogenetic relationships are not known for the included taxa, which prevents a more nuanced approach. Only two families in this study were diverse enough to be treated as separate factors: Trachyleberididae (56 species) and Cytherideidae (12 species). All remaining families were lumped together as a background family rate (25 species). We considered models in which speciation, extinction and preservation probabilities depended on occupancy and taxonomic family individually and combined. Because sexual dimorphism, occupancy and family membership were all individually predictive of extinction risk, we also fit additional models in which extinction depended on these variables in combination. Finally, we also assessed whether sexual size dimorphism was more predictive of extinction when computed as the absolute value of the size difference between sexes, rather than the signed difference, male minus female, as described above. Model support was modestly lower ($\Delta\text{AICc} \approx 1$) when using absolute size differences, indicating that our results are more consistent with extinction risk being influenced by male reproductive investment rather than the absolute size difference between the sexes.

In total, we fit 576 different model configurations: 12 extinction models \times 8 speciation models \times 6 preservation models. We used ΔAICc and Akaike weights to summarize model support, and used model averaging²⁹ to compute coefficient estimates and confidence limits that account for model uncertainty. Continuous covariates were related to probabilities through logit link functions and differing time spans between samples were accounted for in the analysis using the ‘time.intervals’ argument in the ‘process.data’ function of RMark. MARK parameterizes models in terms of survivorship, rather than extinction. To present the results more intuitively, we computed extinction probabilities as $1 - \text{survivorship probabilities}$, and reversed the signs of coefficients so that higher, positive numbers reflecting increasing extinction risk. We also converted probabilities of extinction over a 1-Myr time span to the more commonly reported extinction rates per Myr using equation A1 from the study by Raup⁴².

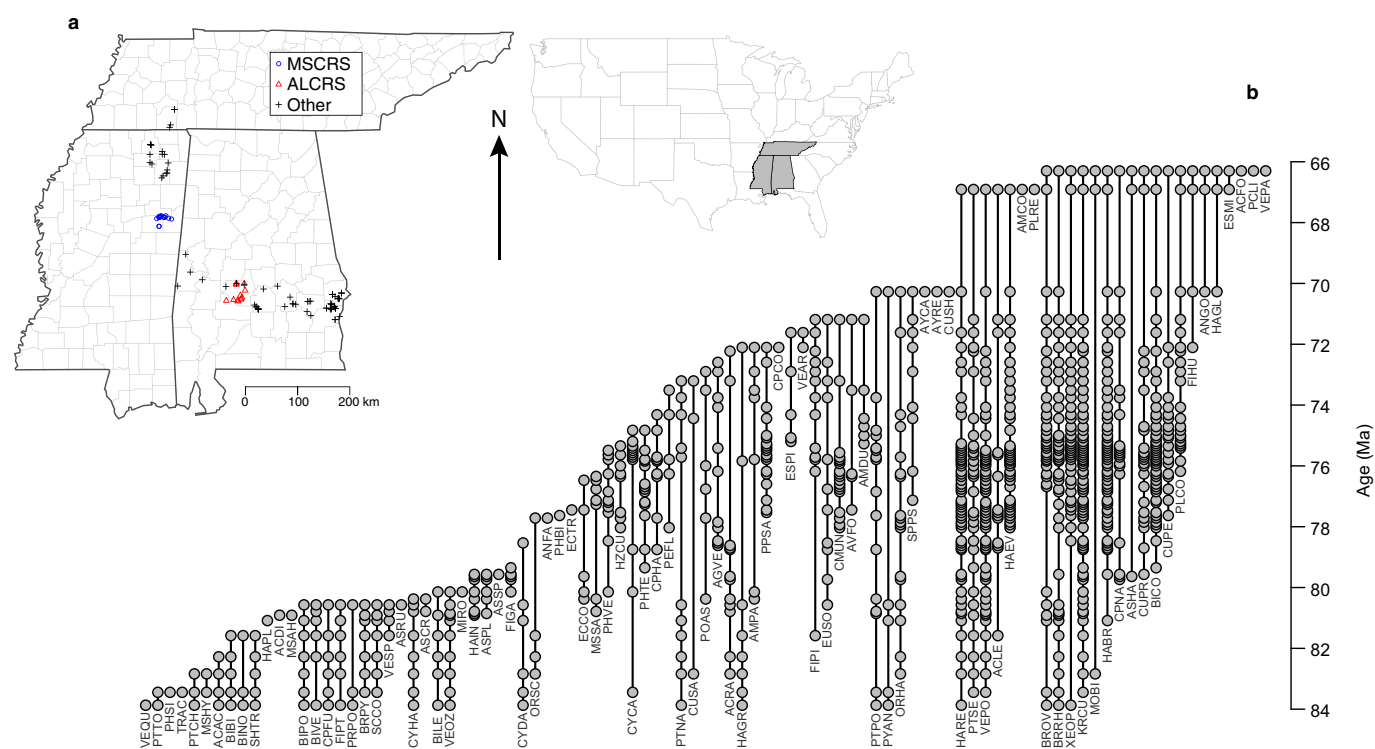
Figure 2 visualizes extinction rates with respect to size and shape dimorphism, as predicted by the model. These predictions were generated by an extinction model that included terms for these two variables, plus occupancy and family membership, using model-averaged coefficients for all terms. Plotted ranges for size and shape dimorphism were chosen to span the values in the observed data. Computing predicted extinction rate from the full model requires values for occupancy and family membership (in addition to size and shape dimorphism). The former was set as the mean occupancy across all species, and the latter, family membership, was set as Trachyleberididae, the most diverse family in the fauna. This figure thus shows predicted extinction for trachyleberidid species with average occupancy, but the patterns discussed are the same under other visualization choices.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The R script to perform the CMR analyses is provided in the Supplementary Information.

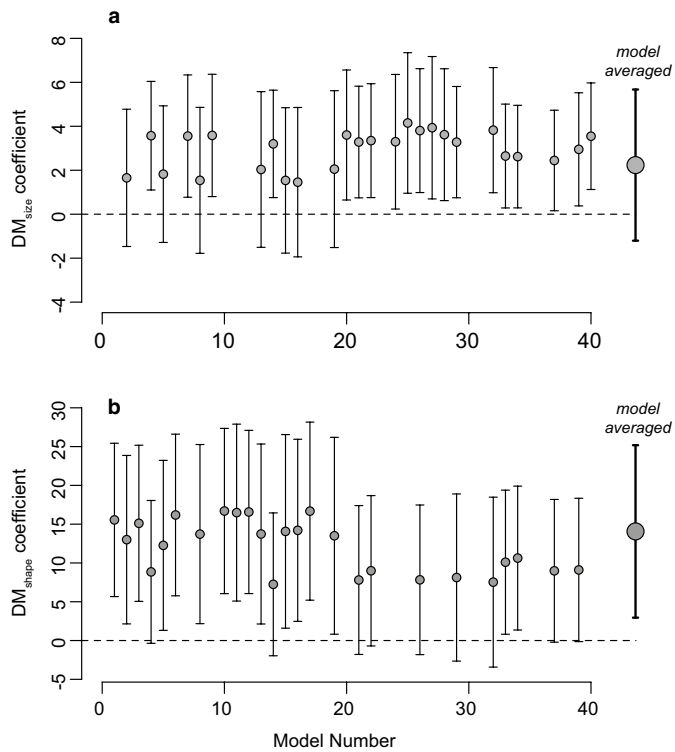
Data availability. Sexual dimorphism data were published previously²⁶ and are shown in Supplementary Table 2. Input files for CMR analyses, which include stratigraphic occurrence (Supplementary Data) and related sample information (Supplementary Table 3), are provided in the Supplementary Information.

37. Larina, E. et al. Upper Maastrichtian ammonite biostratigraphy of the Gulf Coastal Plain (Mississippi Embayment, southern USA). *Cretac. Res.* **60**, 128–151 (2016).
38. White, G. C. & Burnham, K. P. Program MARK: survival estimation from populations of marked animals. *Bird Study* **46**, S120–S139 (1999).
39. Laake, J. L. *RMark: An R Interface for Analysis of Capture-Recapture Data with MARK*. AFSC Processed Report No. 2013-01 (Alaska Fisheries Science Center, 2013).
40. Liow, L. H. & Finarelli, J. A. A dynamic global equilibrium in carnivoran diversification over 20 million years. *Proc. R. Soc. B* **281**, 20132312 (2014).
41. Foote, M. et al. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science* **318**, 1131–1134 (2007).
42. Raup, D. M. Mathematical models of cladogenesis. *Paleobiology* **11**, 42–52 (1985).

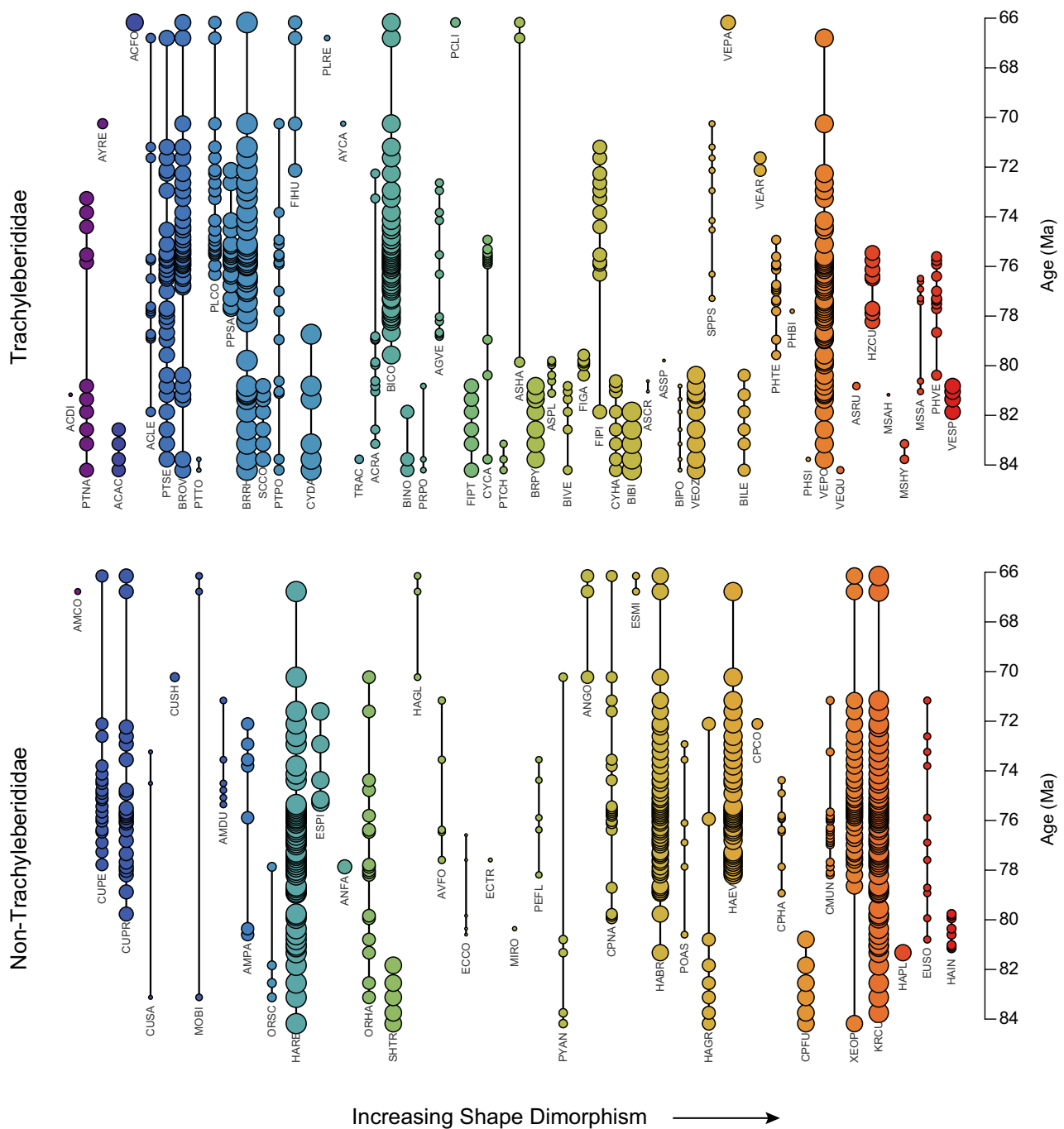


Extended Data Fig. 1 | Stratigraphic section showing the occurrence of 93 species over time. a, Location map of Tennessee, Mississippi and Alabama. The locations of samples that were collected from the focal composite reference section in Mississippi (MSCRS, blue circles) and the composite section in Alabama, which were treated as a replicate (ALCRS, red triangles), are shown along with the additional samples in the

database that were used to compute occupancy (crosses). **b**, Stratigraphic occurrences for the MSCRS are shown. Each grey circle represents the occurrence of a species in a sample, with each species labelled according to four-letter abbreviations given in Supplementary Table 4. The map was made using the R package ‘maps’.



Extended Data Fig. 2 | Estimated model coefficients relating sexual size and shape dimorphism to extinction. a, Sexual size dimorphism (DM_{size}). b, Shape dimorphism (DM_{shape}). The best 40 models are shown, sorted in order of decreasing support. The model-averaged coefficients are shown on the far right as larger circles. These estimates integrate over all models, weighted by their support, appropriately accounting for uncertainty in model selection. Error bars are 95% confidence intervals generated by MARK software.



Extended Data Fig. 3 | Stratigraphic occurrences of species plotted with respect to shape dimorphism. Top, species in the family Trachyleberididae; bottom, all other species. Species are sorted left to right based on shape dimorphism, with more extreme dimorphism plotted

towards the right and in warmer colours. Symbol size is proportional to occupancy (larger indicates more broadly distributed). In the Trachyleberididae, there is a clear visual indication that more strongly dimorphic species have shorter stratigraphic durations.

Extended Data Table 1 | Best supported models for extinction and speciation using occurrence data from a replicate reference section in central Alabama

| Rank | Extinction | Speciation | ΔAICc |
|------|---|------------|---------------------|
| 1 | DM _{size} + DM _{shape} + Occupancy | time | 0.00 |
| 2 | DM _{shape} + Occupancy + Family | time | 3.27 |
| 3 | DM _{size} + Occupancy + Family | time | 3.57 |
| 4 | DM _{size} + DM _{shape} + Occupancy + Family | time | 3.73 |
| 5 | Occupancy | time | 4.88 |
| 6 | Occupancy + Family | time | 5.28 |
| 7 | DM _{size} + DM _{shape} + Occupancy | time | 5.29 |
| 8 | DM _{size} | time | 6.48 |
| 9 | DM _{size} + DM _{shape} | time | 6.96 |
| 10 | DM _{shape} | time | 7.85 |
| 11 | DM _{shape} + Occupancy + Family | time | 8.39 |
| 12 | DM _{size} + Occupancy + Family | time | 8.64 |
| 13 | DM _{size} + DM _{shape} + Occupancy + Family | time | 8.78 |
| 14 | constant | time | 9.24 |
| 15 | Family | time | 9.51 |

Models are listed in order of decreasing model support as measured by ΔAICc ; all models with non-negligible support ($\Delta\text{AICc} < 10$) are shown. The next two columns list the covariates that influence extinction and speciation, respectively, under each model. DM_{size}, sexual size dimorphism; DM_{shape}, sexual shape dimorphism. Occupancy measures how widespread a species is and family indicates the taxonomic family. Constant indicates that speciation or extinction probabilities are the same in each time interval and do not depend on sexual dimorphism or any other covariate. Similar to the Mississippi reference section, the best supported model in which extinction does not depend on sexual dimorphism (model 5) has substantially less support than the best model. Preservation probabilities for these models are a function of occupancy, and in some cases, also include a factor for geological formation/member.

Aspm knockout ferret reveals an evolutionary mechanism governing cerebral cortical size

Matthew B. Johnson^{1,2}, Xingshen Sun^{3,4,5,15}, Andrew Kodani^{1,2,15}, Rebeca Borges-Monroy^{1,2,15}, Kelly M. Girsakis^{1,2}, Steven C. Ryu^{1,2}, Peter P. Wang^{1,2}, Komal Patel⁶, Dilenny M. Gonzalez^{1,2}, Yu Mi Woo⁷, Ziyang Yan^{3,4,5}, Bo Liang^{3,4,5}, Richard S. Smith^{1,2}, Manavi Chatterjee⁶, Daniel Coman^{8,9,10}, Xenophon Papademetris^{9,10,11}, Lawrence H. Staib^{10,11,12}, Fahmeed Hyder^{8,9,10,11}, Joseph B. Mandeville¹³, P. Ellen Grant¹⁴, Kiho Im¹⁴, Hojoong Kwak⁷, John F. Engelhardt^{3,4,5}, Christopher A. Walsh^{1,2*} & Byoung-Il Bae^{1,2,6*}

The human cerebral cortex is distinguished by its large size and abundant gyrification, or folding. However, the evolutionary mechanisms that drive cortical size and structure are unknown. Although genes that are essential for cortical developmental expansion have been identified from the genetics of human primary microcephaly (a disorder associated with reduced brain size and intellectual disability)¹, studies of these genes in mice, which have a smooth cortex that is one thousand times smaller than the cortex of humans, have provided limited insight. Mutations in abnormal spindle-like microcephaly-associated (ASPM), the most common recessive microcephaly gene, reduce cortical volume by at least 50% in humans^{2–4}, but have little effect on the brains of mice^{5–9}; this probably reflects evolutionarily divergent functions of ASPM^{10,11}. Here we used genome editing to create a germline knockout of *Aspm* in the ferret (*Mustela putorius furo*), a species with a larger, gyrified cortex and greater neural progenitor cell diversity^{12–14} than mice, and closer protein sequence homology to the human ASPM protein. *Aspm* knockout ferrets exhibit severe microcephaly (25–40% decreases in brain weight), reflecting reduced cortical surface area without significant change in cortical thickness, as has been found in human patients^{3,4}, suggesting that loss of ‘cortical units’ has occurred. The cortex of fetal *Aspm* knockout ferrets displays a very large premature displacement of ventricular radial glial cells to the outer subventricular zone, where many resemble outer radial glia, a subtype of neural progenitor cells that are essentially absent in mice and have been implicated in cerebral cortical expansion in primates^{12–16}. These data suggest an evolutionary mechanism by which ASPM regulates cortical expansion by controlling the affinity of ventricular radial glial cells for the ventricular surface, thus modulating the ratio of ventricular radial glial cells, the most undifferentiated cell type, to outer radial glia, a more differentiated progenitor.

We injected 148 ferret zygotes with genome editing constructs that targeted *Aspm* exon 15, mutations in which cause severe microcephaly in humans¹⁷, and recovered 11 kits born at full term, all carrying insertions or deletions in the targeted exon (Fig. 1a–d). We established three stable *Aspm* germline knockout ferret lines, which showed comparable phenotypes. Loss of ASPM protein was confirmed in embryonic fibroblasts (Fig. 1e).

Aspm knockout ferrets displayed robust microcephaly (Fig. 1f–i), with up to 40% reduced brain weight (Fig. 1n), but no change in body weight (Fig. 1p), closely modelling the effects of human mutations^{2–4,17}. Magnetic resonance imaging (MRI) showed that, similar to humans⁴,

loss of cortical volume and surface area followed an anterior-to-posterior gradient, and the frontal cortex was affected the most (Fig. 1f–k and Extended Data Table 1). By contrast, the thickness of the cortex in knockout ferrets was preserved, similar to the cortex of human patients with mutations in *ASPM*^{2–4}, and the cytoarchitecture and lamination of neurons was also preserved (Fig. 1l, m, o and Extended Data Fig. 1). This phenotype is distinct from *Aspm* knockout mice, which show approximately 10% reduced brain weight, variable body weight reduction, variable cortical thinning, and no discernible change in cortical surface area^{5–9} (Fig. 1q). Therefore, the *Aspm* loss-of-function phenotype is more similar in ferrets than in mice to the phenotype of human patients with mutations in *ASPM*.

To elucidate the developmental mechanism by which microcephaly occurs, we analysed *Aspm* knockout ferrets during cortical neurogenesis (Fig. 2a–o and Extended Data Figs. 2, 3), which begins around embryonic day 24 (E24) and continues for two weeks after birth, at E41. In the embryonic cortex of wild-type ferrets, undifferentiated ventricular radial glial cells (VRGs) divide symmetrically to expand the pool of VRGs or divide asymmetrically to produce two distinct, more differentiated progenitor subtypes, intermediate progenitors and outer radial glia (ORGs; Fig. 1a). ORGs are multipotent, proliferative, unipolar progenitors, which are abundant in the outer subventricular zone (OSVZ), that express molecular markers, which are also expressed by VRGs, including SOX2, PAX6 and vimentin (VIM); whereas intermediate progenitors are neuronally fated, multipolar transit amplifying cells that predominate in the inner subventricular zone (ISVZ) and express TBR2 (which is encoded by *Eomes*)^{12–14,16}. All three neural progenitor cell (NPC) populations express the mitotic marker Ki-67 and produce neurons that migrate radially into the cortical plate^{12–14,16,19}. The cortex of *Aspm*^{+/-} ferrets at E35 and postnatal day 0 (P0) displayed a ventricular zone that was densely packed with PAX6⁺ or SOX2⁺ VRGs, and a less-dense zone of Ki-67⁺ NPCs that expressed SOX2, TBR2 or both in the SVZ (Fig. 2d–g and Extended Data Figs. 2, 3). By contrast, the cortex of *Aspm* knockout ferrets contained overabundant Ki-67⁺ NPCs in the basal SVZ and intermediate zone (Fig. 2e, f), reminiscent of the positioning of ORGs that normally populate the OSVZ^{12–14,20}. Discontinuous clusters of basal NPCs were accompanied by thinning of the ventricular zone, suggesting that precocious OSVZ progenitors were derived by premature withdrawal from the ventricular zone (Fig. 2d–f and Extended Data Figs. 2, 3). Displaced OSVZ progenitors were more abundant frontally and dorsally (Fig. 2a–c), matching the topography of cortical volume reduction in the adult (Fig. 1f–k).

¹Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children’s Hospital, Harvard Medical School, Boston, MA, USA. ²Howard Hughes Medical Institute, Boston Children’s Hospital, Harvard Medical School, Boston, MA, USA. ³Department of Anatomy and Cell Biology, Center for Gene Therapy, University of Iowa, Iowa City, IA, USA. ⁴Center for Gene Therapy, University of Iowa, Iowa City, IA, USA. ⁵National Ferret Resource and Research Center, University of Iowa, Iowa City, IA, USA. ⁶Department of Neurosurgery, School of Medicine, Yale University, New Haven, CT, USA. ⁷Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA. ⁸Magnetic Resonance Research Center (MRRC), Yale University, New Haven, CT, USA. ⁹Quantitative Neuroscience with Magnetic Resonance (QNMR) Core Center, Yale University, New Haven, CT, USA. ¹⁰Department of Radiology & Biomedical Imaging, Yale University, New Haven, CT, USA. ¹¹Department of Biomedical Engineering, Yale University, New Haven, CT, USA. ¹²Department of Electrical Engineering, Yale University, New Haven, CT, USA. ¹³Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA. ¹⁴Division of Newborn Medicine, Fetal Neonatal Neuroimaging and Developmental Science Center, Boston Children’s Hospital, Harvard Medical School, Boston, MA, USA. ¹⁵These authors contributed equally: Xingshen Sun, Andrew Kodani, Rebeca Borges-Monroy. *e-mail: christopher.walsh@childrens.harvard.edu; byoung-il.bae@yale.edu

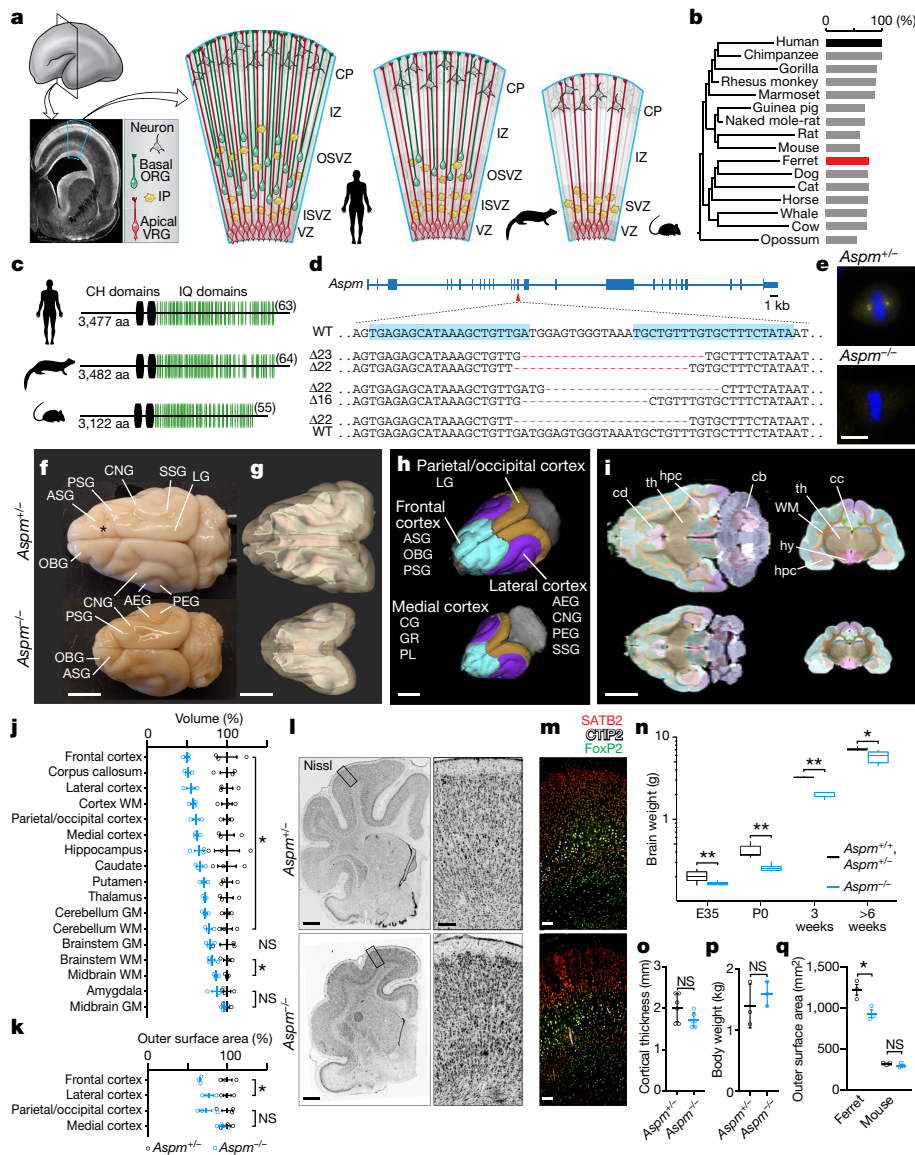


Fig. 1 | *Aspm* knockout ferrets robustly model human microcephaly. **a**, NPC diversity in humans, ferrets and mice. **b**, **c**, ASPM protein is highly similar between humans and ferrets, including the number of calmodulin-binding (IQ) domains (**c**, in parentheses). **d**, Ferret *Aspm* gene showing the targeted sequences (blue highlights) and founder frameshift deletions. **e**, Loss of *Aspm* in knockout ferret embryonic fibroblasts. **f**, Brains of *Aspm*^{+/+} and *Aspm*^{-/-} ferret littermates. **g–k**, MRI segmentations of grey and white matter (**g**), gyri grouped into four regions (**h**), horizontal and coronal sections (**i**) and quantification of volume (**j**) and cortical surface area (**k**). **P* < 0.05; *n* = 3 ferrets per genotype. **l–p**, *Aspm*^{-/-} ferrets show reduced brain weight (**n**, ***P* < 0.005; **P* < 0.01; *n* = 3–17 ferrets per genotype per age group), but cytoarchitecture (**l**), laminar organization (**m**), cortical thickness (**o**, *n* = 6 ferrets per genotype) and body weight (**p**, *n* = 3 ferrets per genotype) are preserved. **q**, Loss of *Aspm* decreases

outer cortical surface area in ferrets, but not in mice (*n* = 3 ferrets and 3 mice per genotype). **P* = 0.0217. Data are mean ± s.e.m. (**j**, **k**, **o–q**); box plots show maximum, third quartile, median, first quartile and minimum (**n**). See Methods, Extended Data Table 1 and Source Data for statistics and reproducibility. Scale bars, 10 μm (**e**), 100 μm (**m**), 1 mm (**l**) and 5 mm (**f–i**). AEG, anterior ectosylvian gyrus; ASG, anterior sigmoid gyrus; cb, cerebellum; cc, corpus callosum; cd, caudate; CG, cingulate gyrus; CH, calponin homology; CNG, coronal gyrus; CP, cortical plate; GR, gyrus rectus; GM, grey matter; hpc, hippocampus; hy, hypothalamus; IZ, intermediate zone; IP, intermediate progenitor; LG, lateral gyrus; OBG, orbital gyrus; PEG, posterior ectosylvian gyrus; PL, piriform lobe; PSG, posterior sigmoid gyrus; SSG, suprasylvian gyrus; th, thalamus; VZ, ventricular zone; WM, white matter.

Many displaced progenitors in the OSVZ of *Aspm* knockout ferrets expressed VRG/ORG markers including VIM, phosphorylated vimentin (pVIM), phosphorylated histone H3 (pH3), SOX2 and PAX6; as well as the ciliary marker ARL13B and the human ORG-enriched genes *Ptprz1* and *Hopx*²¹, whereas other cells expressed the intermediate progenitor marker, as shown by TBR2 protein and *Eomes* mRNA analysis (Fig. 2g–o and Extended Data Figs. 2, 3, 6). Some of the displaced cells had an ORG-like unipolar morphology, with basal radial fibres that were immunoreactive to VIM, pVIM or HOPX antibodies (Fig. 2g–i, k, o). Quantification of pVIM⁺ mitotic NPCs revealed a threefold increase in the number of ORG-like progenitors in the knockout ferrets at E35

(*P* = 0.006; 3 *Aspm*^{+/+} and 4 *Aspm*^{-/-} littermates; Fig. 2j). The intermingled presence of NEUROG2⁺HOPX⁺ ORGs, TBR2⁺ intermediate progenitors and DCX⁺ newborn neurons together indicated preserved neurogenesis within clusters of displaced NPCs (Fig. 2n and Extended Data Fig. 6). These data demonstrate that the loss of *Aspm* in the ferret cortex causes VRGs to prematurely detach from the ventricular zone and relocate to the OSVZ, where many dislocated cells exhibit ORG morphology, molecular profile and neurogenic potential.

These marked changes in NPC populations in the knockout ferret contrast with six previously reported *Aspm* knockout mouse lines^{5–9}, which have consistently shown limited changes in NPC identity and

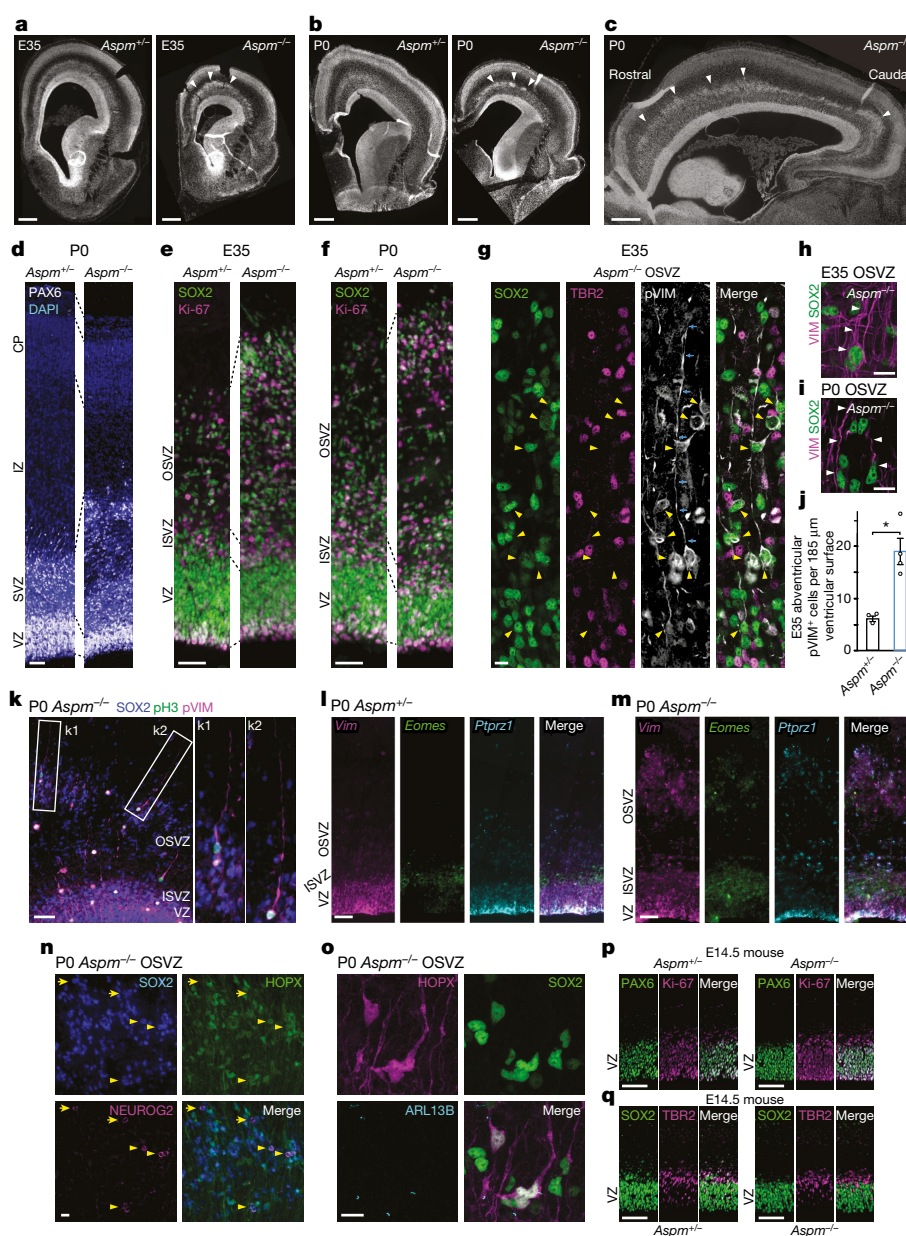


Fig. 2 | *Aspm* knockout ferrets show displaced NPCs. **a–f**, Nuclear staining of *Aspm*^{−/−} ferret brains shows that there is a premature OSVZ-like zone (**a–c**, arrowheads), which contains NPCs that express PAX6, SOX2, and Ki-67 (**d–f**). **g–k**, Displaced NPCs include SOX2⁺pVIM⁺ ORG (**g**, arrowheads) with a basal process (**g**, arrows; **h**, **i**, **k**) and TBR2⁺ intermediate progenitors. The number of adventitious pVIM⁺ NPCs are increased threefold in *Aspm*^{−/−} ferrets (**j**). **P* = 0.006; analysed using

a one-tailed *t*-test; data are mean ± s.e.m.; *n* = 3 *Aspm*^{+/+} and *n* = 4 *Aspm*^{−/−} animals. **l–o**, Displaced NPCs express *Vim*, *Eomes* or *Ptpz1* (**l**, **m**), have ARL13B⁺ cilia (**o**) and are either SOX2⁺NEUROG2⁺HOPX⁺ (arrowheads) or SOX2⁺NEUROG2⁺HOPX[−] (arrows) (**n**). **p**, **q**, *Aspm*^{−/−} mice lack displaced NPCs. See Methods for statistics and reproducibility. Scale bars, 10 μm (**g–i**, **n**, **o**), 50 μm (**d–f**), 100 μm (**k–m**, **p**, **q**) and 500 μm (**a–c**).

organization. *Aspm* knockout mice show a trend towards an increased number of intermediate progenitors at the expense of VRGs⁹, but lack ectopic basal SOX2⁺ or PAX6⁺ NPCs (Fig. 2p, q and Extended Data Fig. 4). *Aspm* knockout ferrets also showed increased cell apoptosis in telencephalic germinal zones that was not seen in *Aspm* knockout mice^{5–7,9} (Extended Data Fig. 5), further highlighting that loss of *Aspm* elicits divergent brain phenotypes in ferrets and mice.

Single-cell RNA-sequencing²² (scRNA-seq) of around 21,000 cells from the telencephalons of seven E35 embryos (3 *Aspm*^{+/+} or *Aspm*^{+/−} and 4 *Aspm*^{−/−} animals) reinforced the conclusion that NPC proportions were altered in the *Aspm* knockout animals, although their transcriptional programs were mostly preserved (Fig. 3, Extended Data Fig. 7 and Extended Data Table 2). We identified cell clusters corresponding to excitatory and inhibitory progenitor and neuronal subtypes, as well as non-neural cells (Fig. 3a, b) and found that the

cell type composition of the E35 *Aspm* knockout forebrain was significantly altered ($\chi^2 = 267.27$, degrees of freedom = 12, $P = 2.2 \times 10^{-16}$), yet cells still clustered by cell type, not by genotype or batch (Extended Data Fig. 7). Consistent with immunohistochemical observations, scRNA-seq analysis suggested that VRGs, wild-type ORGs and prematurely displaced knockout ORGs were transcriptionally indistinguishable, and the total proportion of radial glial cells (cycling radial glial cells and interphase radial glial cells) was unchanged in *Aspm* knockout cells (Extended Data Table 2). A 30% increase in the proportion of intermediate progenitors in knockout ferrets ($P = 0.0002$, false discovery rate of < 0.01; Fig. 3c and Extended Data Table 2) was consistent with the increased number of intermediate progenitors that were detected by immunostaining (Fig. 2l, m and Extended Data Fig. 6) and was further validated by single-molecule fluorescence in situ hybridization (Fig. 3d, e). A doubling of the small proportion of

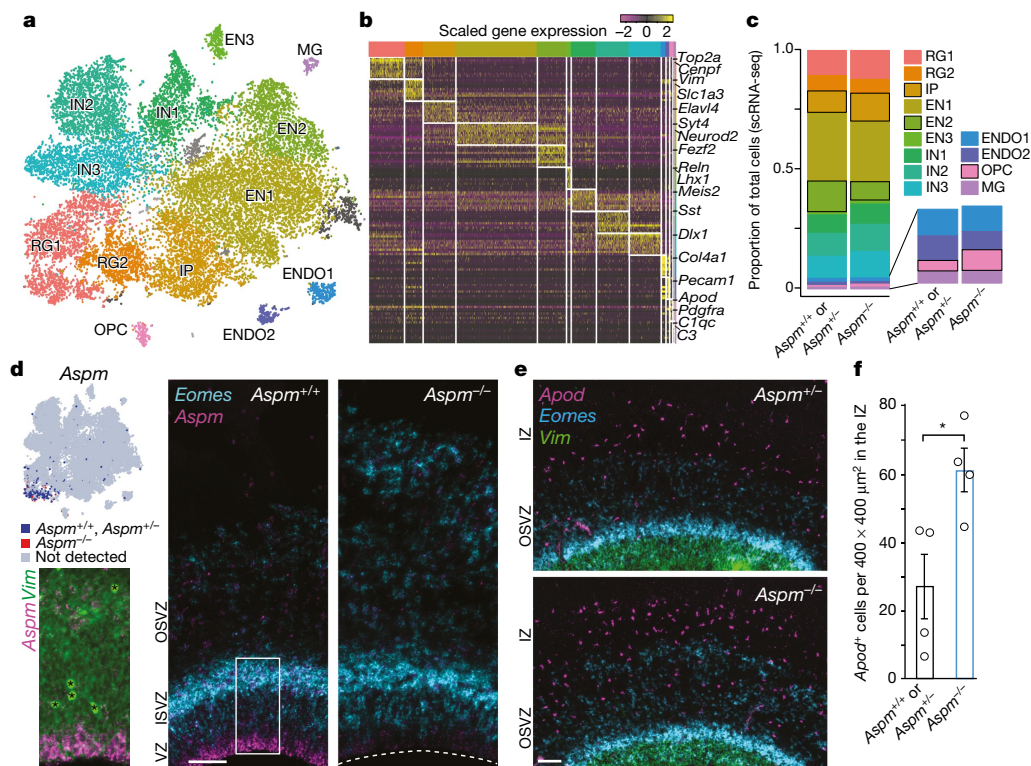


Fig. 3 | Loss of *Aspm* changes cell type proportions but not transcriptional programs. **a**, scRNA-seq identifies major cell types at E35. For abbreviations and statistics, see Extended Data Table 2. **b**, Cell type markers for each cluster. **c**, Proportions of each cell type, with the largest changes indicated by black outlines (bootstrap false discovery rate < 0.01). **d**, *Aspm* is enriched at the apical surface of the ventricular zone, whereas

the number of *Eomes*⁺ intermediate progenitors is increased in the SVZ of *Aspm* knockout ferrets. Asterisks indicate blood vessels. **e**, **f**, The number of *Apod*⁺ oligodendrocyte precursor cells is increased in *Aspm*^{-/-} ferrets. **P* = 0.012; analysed using a one-tailed *t*-test; *n* = 4 animals per genotype. See Methods for statistics and reproducibility. Data are mean ± s.e.m. Scale bars, 100 μm.

cells that expressed the oligodendrocyte precursor cell markers *Apod* and *Olig1* suggested limited but significant premature differentiation towards the glial lineage (Fig. 3c, e, f and Extended Data Table 2). These scRNA-seq data suggest that the gene expression programs of cortical neurogenesis are mostly preserved, but the proportions of developmental cell types are changed, in microcephaly associated with ASPM mutations.

While examining potential molecular mechanisms for the detachment of VRGs from the ventricular surface in the knockout ferret, we identified a novel interaction between ASPM, which is localized at the centrosome, and the apical polarity complex (Fig. 4). Together with other centrosomal proteins, ASPM is essential for normal stem cell behaviour, such as centriole biogenesis and maternal centriole structure (Extended Data Fig. 8), and interactions between the mother centriole and the apical membrane have been implicated in the maintenance of NPC stem cell character^{9,23–25}. In VRGs, the centrosome is localized to the ventricular endfeet, which are linked by adherens junctions to form a polarized neuroepithelium, which expresses apical polarity complex proteins at the ventricular surface²⁶. *Aspm*^{-/-} mice showed abrogated staining of the core apical complex protein aPKCζ along the disrupted ventricular surface at E14.5 (Fig. 4a). Intriguingly, we found that depletion of ASPM by RNA interference in H4 human neuroglioma cells resulted in the loss of PKCζ and another critical apical complex protein, PAR6α, from the centrosome (Fig. 4b). Furthermore, we found an interaction between ASPM and PKCζ, as indicated by mutual co-immunoprecipitation (Fig. 4c), that may mediate centrosomal recruitment of the apical polarity complex, providing a new mechanistic insight into the link between centrosomal microcephaly-related proteins and apical progenitor identity.

Finally, we found sharply reduced staining for ninein, another microcephaly-associated centrosomal protein²⁷, at both the E14.5 mouse and E35 ferret ventricular surface (Fig. 4d, e). Ninein localizes

to the mother centriole and is critical for NPC maintenance^{25,28}, and depletion of either ASPM or PKCζ in H4 cells reduced centrosomal localization of ninein (Fig. 4f). Importantly, *Aspm*^{-/-} mouse embryonic fibroblasts expressed normal levels of aPKCζ, PAR6, and ninein proteins, suggesting that loss of aPKCζ and ninein from the ventricular surface of *Aspm* knockout mice and ferrets is primarily because of mislocalization, rather than downregulation. These data show that loss of ASPM disturbs the organization and function of the centrosome at multiple levels, and suggest disruption of the centrosome–apical polarity complex interface as a mechanism underlying the displacement of VRGs from the ventricular zone in the *Aspm* knockout ferret.

Collectively, our data show that ASPM regulates the affinity of VRGs for the ventricular surface. Displaced mutant progenitors show many features of ORGs, indicating that ASPM has a central role in the regulation of the normal timing of the transition from VRG to ORG, and thus the ratio of VRGs to ORGs over the course of development. Premature basal displacement deprives VRGs of proliferation-inducing factors obtained from the cerebrospinal fluid²⁹, increases the proportions of less-proliferative ORGs and intermediate progenitors, and results in a smaller cerebral cortex. The frontal predominance that characterizes both the loss of cortical surface area and VRG displacement further indicates that the premature transformation of VRG to ORG leads directly to reduced cortical units and surface area.

Our results support the idea that expansion of cortical surface area during human evolution may have arisen in part from changes to the proliferative time window of VRGs. Changes in the amino acid sequence of ASPM and other microcephaly-associated centrosomal proteins¹⁰ may have affected the timing of the VRG proliferative window by altering interactions between maternal centriole components and the apical polarity complex. Finally, we find that for human brain disorders that are poorly recapitulated in the mouse or in cell culture, the ferret is an efficient and accurate genetic model that demonstrates

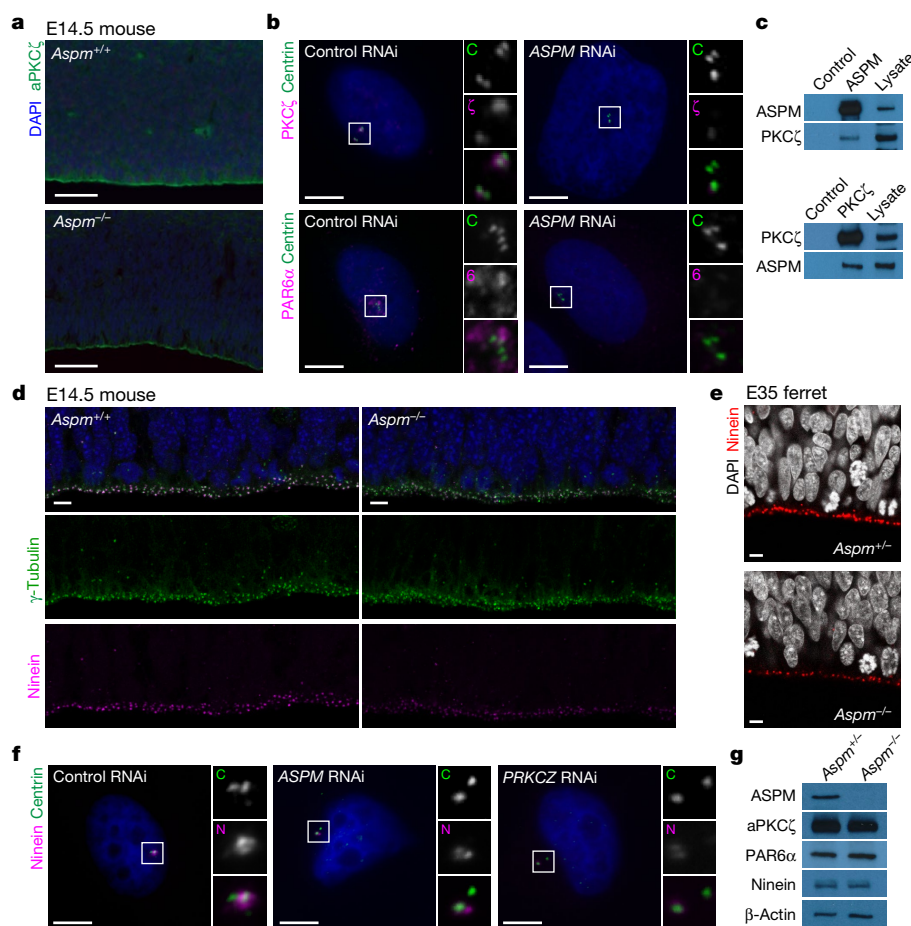


Fig. 4 | ASPM controls localization of apical polarity complex proteins to the centrosome. **a**, aPKC ζ at the ventricular surface is decreased in *Aspm*^{-/-} mice. **b**, Depletion of ASPM in H4 cells prevents recruitment of PKC ζ and PAR6 α to the centrosomes. RNAi, RNA interference. **c**, ASPM and PKC ζ co-immunoprecipitate in extracts of HeLa cells. **d**, **e**, Loss of ASPM decreases ventricular surface staining for ninein in mice (**d**) and

ferrets (**e**). **f**, Depletion of ASPM and PKC ζ prevents recruitment of ninein to the centrosomes in H4 cells. **g**, Levels of aPKC ζ , PAR6 α , and ninein are unchanged in *Aspm*^{-/-} mouse embryonic fibroblasts. See Methods for reproducibility. For gel source data, see Supplementary Data. Scale bars, 5 μ m (**b**, **d**-**f**) and 50 μ m (**a**).

robust phenotypes and can be used to investigate the mechanisms underlying disorders of the brain.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0035-0>.

Received: 25 May 2017; Accepted: 22 February 2018;

Published online: 11 April 2018

1. Faheem, M. et al. Molecular genetics of human primary microcephaly: an overview. *BMC Med. Genomics* **8**, S4 (2015).
2. Bond, J. et al. *ASPM* is a major determinant of cerebral cortical size. *Nat. Genet.* **32**, 316–320 (2002).
3. Passemar, S. et al. Abnormal spindle-like microcephaly-associated (*ASPM*) mutations strongly disrupt neocortical structure but spare the hippocampus and long-term memory. *Cortex* **74**, 158–176 (2016).
4. Desir, J., Cassart, M., David, P., Van Bogaert, P. & Abramowicz, M. Primary microcephaly with *ASPM* mutation shows simplified cortical gyration with antero-posterior gradient pre- and post-natally. *Am. J. Med. Genet. A* **146A**, 1439–1443 (2008).
5. Pulvers, J. N. et al. Mutations in mouse *Aspm* (abnormal spindle-like microcephaly associated) cause not only microcephaly but also major defects in the germline. *Proc. Natl Acad. Sci. USA* **107**, 16595–16600 (2010).
6. Fujimori, A. et al. Disruption of *Aspm* causes microcephaly with abnormal neuronal differentiation. *Brain Dev.* **36**, 661–669 (2014).
7. Capecci, M. R. & Pozner, A. *ASPM* regulates symmetric stem cell division by tuning cyclin E ubiquitination. *Nat. Commun.* **6**, 8763 (2015).
8. Williams, S. E. et al. *Aspm* sustains postnatal cerebellar neurogenesis and medulloblastoma growth in mice. *Development* **142**, 3921–3932 (2015).

9. Jayaraman, D. et al. Microcephaly proteins Wdr62 and *Aspm* define a mother centriole complex regulating centriole biogenesis, apical complex, and cell fate. *Neuron* **92**, 813–828 (2016).
10. Montgomery, S. H. & Mundy, N. I. Microcephaly genes evolved adaptively throughout the evolution of eutherian mammals. *BMC Evol. Biol.* **14**, 120 (2014).
11. Bae, B. I., Jayaraman, D. & Walsh, C. A. Genetic changes shaping the human brain. *Dev. Cell* **32**, 423–434 (2015).
12. Fietz, S. A. et al. OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* **13**, 690–699 (2010).
13. Reillo, I., de Juan Romero, C., García-Cabezas, M. A. & Borrell, V. A role for intermediate radial glia in the tangential expansion of the mammalian cerebral cortex. *Cereb. Cortex* **21**, 1674–1694 (2011).
14. Johnson, M. B. et al. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat. Neurosci.* **18**, 637–646 (2015).
15. Smart, I. H., Dehay, C., Giroud, P., Berland, M. & Kennedy, H. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb. Cortex* **12**, 37–53 (2002).
16. Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
17. Bond, J. et al. Protein-truncating mutations in *ASPM* cause variable reduction in brain size. *Am. J. Hum. Genet.* **73**, 1170–1177 (2003).
18. Hutchinson, E. B. et al. Population based MRI and DTI templates of the adult ferret brain and tools for voxelwise analysis. *Neuroimage* **152**, 575–589 (2017).
19. Martínez-Cerdeño, V. et al. Comparative analysis of the subventricular zone in rat, ferret and macaque: evidence for an outer subventricular zone in rodents. *PLoS ONE* **7**, e30178 (2012).
20. Martínez-Martínez, M. A. et al. A restricted period for formation of outer subventricular zone defined by *Cdh1* and *Trnp1* levels. *Nat. Commun.* **7**, 11812 (2016).
21. Pollen, A. A. et al. Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

22. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
23. Kodani, A. et al. Centriolar satellites assemble centrosomal microcephaly proteins to recruit CDK2 and promote centriole duplication. *eLife* **4**, e07519 (2015).
24. Paridaen, J. T., Wilsch-Bräuninger, M. & Huttner, W. B. Asymmetric inheritance of centrosome-associated primary cilium membrane directs ciliogenesis after cell division. *Cell* **155**, 333–344 (2013).
25. Wang, X. et al. Asymmetric centrosome inheritance maintains neural progenitors in the neocortex. *Nature* **461**, 947–955 (2009).
26. Singh, S. & Solecki, D. J. Polarity transitions during neurogenesis and germinal zone exit in the developing central nervous system. *Front. Cell. Neurosci.* **9**, 62 (2015).
27. Dauber, A. et al. Novel microcephalic primordial dwarfism disorder associated with variants in the centrosomal protein ninein. *J. Clin. Endocrinol. Metab.* **97**, E2140–E2151 (2012).
28. Zhang, X. et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* **166**, 1147–1162 (2016).
29. Lehtinen, M. K. et al. The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. *Neuron* **69**, 893–905 (2011).

Acknowledgements We thank the late R. W. Guillery, who first introduced ferrets as a model for developmental neuroscience; J. K. Joung for advice on genome editing; J. Bond for the ASPM antibody; L. Vasung, P. Herman, J. Neil and C. D. Kroenke for advice on ferret brain MRI; A. Lee, the G. M. Church laboratory (S. Biwas), the S. McCarroll laboratory (S. Burger), the P. Kharchenko laboratory (J. Fan), and the R. Satija laboratory (A. Butler) for advice on scRNA-seq; S. Wasiuk, E. Feiner, A. S. Kamumbu and M. Lee for technical assistance; Marshall BioResources for animal husbandry; and E. Pollack and the veterinary staff at Boston Children's Hospital and Yale School of Medicine for surgical support. Animal silhouettes in Fig. 1 were designed by Freepik from <https://www.flaticon.com/>. This work was supported by P30NS052519 (F.H. and Yale's QNMR Core Center), 2R01MH067528 (F.H.), 1R24MH114805 (X.P.), R21HD083956

(K.I.), R01EB017337 (P.E.G.), R24HL123482 (J.F.E.), 5R01NS032457 (C.A.W.), 5R21NS091865 (B.-I.B.) and the Allen Discovery Center program through The Paul G. Allen Frontiers Group. C.A.W. is an Investigator of the Howard Hughes Medical Institute.

Reviewer information *Nature* thanks S. Juliano, F. Tissir and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions B.-I.B., C.A.W. and J.F.E. conceived the project. B.-I.B. generated genome-editing reagents. X.S., Z.Y. and B.L. injected ferret zygotes under J.F.E.'s supervision. M.B.J., K.M.G., P.P.W. and D.M.G. performed immunohistochemistry experiments. R.S.S., M.C., K.I., J.B.M., P.E.G., D.C., X.P., L.H.S. and F.H. performed MRI analyses. M.B.J., R.B.-M., Y.M.W. and H.K. performed scRNA-seq. M.B.J. and R.B.-M. performed single-molecule fluorescence in situ hybridization. A.K. and S.C.R. characterized molecular defects with K.P. and B.-I.B. With input from all authors, M.B.J., B.-I.B. and C.A.W. interpreted the data and wrote the paper.

Competing interests X.P. is a consultant for Electrical Geodesics Inc. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0035-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0035-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to C.A.W. or B.-I.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

We complied with all relevant ethical regulations and the experiments that we performed were approved by the Institutional Animal Care and Use Committees (IACUCs) at the University of Iowa, Boston Children's Hospital, Yale School of Medicine and Marshall BioResources.

ASPM protein homology and domain analysis. ASPM protein sequences of 16 mammals were extracted from NCBI GenBank. Global pairwise alignment was performed with EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The percentage of homology to human ASPM was calculated for each animal based on the alignment score using the Needleman–Wunsch algorithm (gap opening penalty, 10; gap extension penalty, 0.5). The phylogenetic tree was generated using http://timetree.org. CH and IQ domains were counted using the simple modular architecture research tool (http://smart.embl.de/), for NP_060606.3 (human), NP_033921.3 (mouse) and ENSMPUT00000010205.1 (ferret).

TALEN assembly and mRNA synthesis. We assembled three pairs of transcription activator-like effector nucleases (TALENs) that target exon 15 of ferret *Aspm*, which encodes the second CH domain, and cloned the TALENs into a mammalian expression vector with CMV and T7 promoters through a commercial service (PNA Bio). Gene targeting efficiency of each TALEN pair was tested in HEK293T cells using a split GFP-based reporter³⁰. The most efficient pair, targeting TGAGAGCATAAAGCTGTGATGGAGTGGGTAAATGCTGTTGTGCTTTCTATA (target–spacer–target) was chosen for genome editing *in vivo*. These plasmids are available from Addgene. For mRNA synthesis, endotoxin-free TALEN plasmids were prepared using NucleoBond Xtra Midi EF kit (Clontech), ethanol-precipitated three times, linearized using ScaI digestion (New England BioLabs) and gel-purified. mRNAs were synthesized using the mMessage mMachine T7 ULTRA kit (ThermoFisher Scientific) and cleaned up using the MEGAclear transcription clean-up kit (ThermoFisher Scientific). Of note, we performed the optional ammonium acetate precipitation to improve the quality of the mRNAs. The TALEN mRNAs were diluted in sterile EmbryoMax injection buffer (Millipore) at 50 ng μ l⁻¹, aliquoted and kept frozen at -150°C until use.

Embryonic targeting of the ferret *Aspm* gene. Zygotes were collected from the mating of ferrets with a sable coat colour (Marshall BioResources) as previously described³¹. TALEN mRNA (50 ng μ l⁻¹) was injected into the cytoplasm of zygotes using a micromanipulator and injector (Eppendorf) and a phase-contrast microscope. Initially 79 ferret zygotes were injected and cultured *in vitro* for five days so that they reached the blastocyst stage. Twelve zygotes developed to blastocysts, from each of these zygotes genomic DNA was extracted and whole-genome multiple displacement amplification was performed. The targeted genomic region was amplified by PCR (primers: 5'-TTTGTGTGTGTTTCAGGTGGA-3' and 5'-TGCATTATACAAGTGGTGACAGA-3' with a 430-bp product size), gel-purified and cloned using a TOPO-TA cloning kit (ThermoFisher Scientific). Twelve plasmid clones were sequenced from individual bacterial colonies from each blastocyst. These studies demonstrated an 87% targeting efficiency (14 insertions or deletions in 16 alleles or 8 blastocysts that we were able to analysis). Next, we injected 148 zygotes, incubated them at 39°C for 24 h, and transferred 116 two-cell-stage embryos into the oviduct of pseudopregnant female sable ferrets as previously described³¹. Twenty-three ferrets were born and eleven survived. All 11 ferrets had insertions or deletions (100% efficiency). The F₀ ferrets suckled and swallowed milk normally and grew without gross abnormalities.

Germline transmission. Six *Aspm* mutant ferrets were shipped to Marshall BioResources at three months of age and maintained according to the protocol approved by IACUC. Two compound heterozygous males, $\Delta 23$; $\Delta 22$ (c.3364_3386del;c.3363_3384del) and $\Delta 22$; $\Delta 16$ (c.3367_3388del;c.3364_3379del), and one heterozygous female, $\Delta 22$; WT (c.3363_3384del; WT), (Fig. 1d) were chosen as founders, because they had similar frameshift, early truncating mutations. They were bred to each other or wild-type ferrets. Germline transmission was confirmed by T7 endonuclease I assay (New England BioLabs) and sequencing of both alleles. Eventually animals with a specific $\Delta 22$ mutation (c.3363_3384del) were maintained for breeding. Routine genotyping was carried out with PCR (primers: 5'-ATCAATAAGAAAAAGACAAAAGAAATAGTGG-3' and 5'-CTTAAGTCAGTGAGCTTAAACAGAAAT-3' with a 150-bp product size for the wild-type allele and 128-bp from the knockout allele). *Aspm* knockout males mated successfully, and knockout kits were born at expected Mendelian ratios.

Semen analysis. Every mating was closely monitored at Marshall BioResources. Sperm samples were collected from mated females directly after mating. The concentration, motility and morphology of the sperm were analysed by an experienced technician. Each male received a sperm check evaluation at least once a month. Samples from *Aspm*^{+/+}, *Aspm*^{+/-} and *Aspm*^{-/-} males showed similar sperm counts.

Ferret colony management and tissue handling. The *Aspm* knockout ferret colony was maintained at Marshall BioResources. For embryonic ages and < P8, timed pregnant jills were shipped to Boston Children's Hospital and euthanized before

embryo extraction, at which point brains were removed from the embryos and drop-fixed in 4% paraformaldehyde (PFA) at 4°C overnight. For \geq P8 ferrets, all animals were deeply anaesthetized and weighed before transcardial perfusion with cold PBS followed by 4% PFA, after which the brains were extracted and placed in 4% PFA at 4°C overnight. The brains were subsequently washed and stored in PBS before processing for immunohistochemistry. All brain weight measurements were made post-fixation, before sucrose infiltration.

MRI. Ferrets. Three *Aspm*^{+/+} and three *Aspm*^{-/-} ferrets (> 8 months of age) were perfused using 4% PFA in PBS. The brains were dissected and post-fixed in PFA and PBS containing 4 mM gadolinium contrast Magnevist (Bayer) at 4°C for two weeks. The brains were scanned using a 9.4 T at Yale Magnetic Resonance Center and the Martinos Center for Biomedical Imaging, Massachusetts General Hospital. At Yale, a custom-made ¹H radiofrequency coil (40-mm diameter) was used for diffusion tensor imaging (DTI). DTI acquisition was obtained with a Stejskal and Tanner spin-echo diffusion-weighted sequence with a diffusion gradient $\delta = 5$ ms and a delay $\Delta = 15$ ms between diffusion gradients³². Sixty-four slices of 500- μ m thickness, field of view of 25.6 \times 25.6 mm² and 128 \times 128 resolution were acquired with a repetition time (TR) of 4 s, echo time (TE) of 30 ms and four averages. Each of the six MR images was first corrected for B1 shading artefacts using a slice inhomogeneity correction³³ and an inverse covariance mapping of grey matter density (D.C., X.P., L.H.S. & F.H., unpublished observations). Next, the Ferret Atlas¹⁸ was registered to each of the MRI images using a tensor *b*-spline normalized mutual information nonlinear intensity-based registration algorithm^{34,35} with a control point spacing of 1 mm. The result of the registration was used to warp the atlas regions to each individual MRI, and from this we calculated the volume of each of the warped regions as shown in Fig. 1 and Extended Data Table 1. Anterior sigmoid gyrus, orbital gyrus and posterior sigmoid gyrus were considered to be frontal cortex; anterior ectosylvian gyrus, coronal gyrus, posterior ectosylvian gyrus, and suprasylvian gyrus were considered to be lateral cortex; cingulate gyrus, gyrus rectus and piriform lobe were considered to be medial cortex; and lateral gyrus was considered to be the parietal/occipital cortex. The name of each brain part in Fig. 1 is as previously described³⁶. For DTI tensor measurement, a total of 15 different non-collinear diffusion weighted directions ($b = 1,000$ s mm⁻²) and 1 without diffusion weighting were obtained. The six elements of the diffusion tensor were calculated from the signal intensity of the diffusion-weighted images. Tensor eigenvalues and their corresponding eigenvectors were computed, along with fractional anisotropy, at each voxel. The images were colour-coded by the principal direction (eigenvector) of diffusion using BioImage Suite³⁷ (http://www.bioimage-suite.org/). At Massachusetts General Hospital, we acquired anatomically accurate brain volume images with minimal distortion using FLASH (fast low angle shot) MRI sequence with TR = 100 ms, TE = 30 ms and 150- μ m isotropic resolution. Cortical grey and white matter were manually segmented using FreeView (http://surfer.nmr.mgh.harvard.edu) and their volumes were measured.

Mice. Three *Aspm*^{+/+} and three *Aspm*^{-/-} mouse brains were dissected from P30 animals perfused with 4% PFA and post-fixed as described above. Brains were submerged into perfluorocarbon oil (Fomblin, Fisher Scientific) at 4°C for three days, and imaged in this oil using a Bruker BioSpec 70/30 7 T MRI scanner (a sub-millimetre MRI with a 30-cm bore and 450 mT m⁻¹ gradient) in the Small Animal Imaging Facility at Boston Children's Hospital. MRI scans were isotropic 63- μ m voxels across the entire brain. Cortical surface area was visualized and measured using the FreeView, Osyrinx and ImageJ 3D projection.

Fluorescent immunohistochemistry. Fixed ferret brains were infiltrated with a series of 10%, then 20% and finally 30% w/v sucrose solutions in PBS until sunk, then embedded in optimal cutting temperature (OCT) compound and frozen in isopentane cooled to -40°C, after which they were stored long-term at -80°C. Brains were sectioned at 10- to 20- μ m thickness on a Leica Cryostat, mounted immediately onto warm charged SuperFrost Plus slides (Fisher Scientific) and dried at 37°C for 10 to 30 min before storage at -80°C. After applying a hydrophobic barrier around the tissue (ImmEdge Pen, Vector Labs), slides were washed in cold 0.1 M PBS followed by antigen retrieval in Retrieval pH 6.0 (BD Biosciences) at 80–90°C in a hybridization oven for 45 min. Sections were then cooled to room temperature in Retrieval, washed in cold 0.1 M PBS, and blocked for 1 h at room temperature (5% normal donkey serum, 1% w/v BSA, 0.2% w/v glycine and 0.2% w/v lysine, in PBS). Slides were incubated with primary antibodies for two nights on a rotary shaker at 4°C in blocking buffer plus 0.3% Triton X-100. Sections were then washed in PBS and incubated for 2 h at room temperature in blocking buffer containing secondary antibodies at 1:500 (Jackson ImmunoResearch). Finally, slides were washed in PBS, counterstained with DAPI at 1 μ g ml⁻¹ in PBS for 15 min, washed again and coverslipped with Fluoromount-G (Southern Biotech). Images were obtained with a Zeiss LSM700 confocal microscope and Leica MZ16 F fluorescence stereomicroscope. The following antibodies were used at 1:200–1:2,000: PAX6 (Abcam ab5790), FoxP2 (Abcam ab16046), CTIP2 (Abcam ab18465), SATB2 (Bethyl A301-864A), SATB2 (Abcam ab51502), SOX2 (SCBT sc-17320), TBR2 (Millipore AB15894), Ki-67 (BD 550609), pVIM (MBL D076-3), VIM (Abcam

ab8978), pH3 (Millipore 07-145), HOPX (SCBT sc-30216), ARL13B (Abcam ab136648), ARL13B (ProteinTech 17711-1-AP), NEUROG2 (R&D MAB3314) and DCX (SCBT sc-8066).

Apoptosis assay. We examined apoptosis on cryosections using the ApopTag Red In situ Apoptosis Detection Kit (Millipore) according to the manufacturer's protocol.

scRNA-seq. Cell capture and sequencing. Cell capture and sequencing were performed using the Drop-seq method²² (<http://mccarrolllab.com/dropseq/>). Forebrain tissue was isolated from 2 *Aspm*^{+/+}, 1 *Aspm*^{+/-} and 4 *Aspm*^{-/-} E35 ferret embryos and cryopreserved³⁸, then shipped to Cornell and processed there for single-cell capture, library preparation and sequencing.

Read alignment and digital gene expression matrix generation. Ferret reference gene annotations were expanded using bulk RNA-seq data from the cortex of two P2 ferrets. Bulk data were first mapped to the Ensembl ferret reference genome and transcriptome using TopHat2, and a transcriptome was assembled with cufflinks; this assembled transcriptome and the Ensembl reference transcriptome version 1.0.85 were merged using cuffmerge³⁹. The Drop-Seq Core Computational Protocol version 1.0.1 was followed²². Fastq reads were converted to BAM using the 'FastqToSam' command in Picard (<http://broadinstitute.github.io/picard/>). Read pairs for which more than one base in the barcode had a quality below 10 were discarded. Adaptor sequences were trimmed from the 5' end of the read, along with polyA tails. Star-2.5.2a⁴⁰ was used to map reads to the custom transcriptome reference. The digital gene expression matrix was extracted using the 'DigitalExpression' program of the Drop-seq protocol²², keeping only cells with at least 200 reads per cell for clustering analysis.

Single cell clustering. Seurat software was used for dimensionality reduction, clustering and obtaining cluster markers⁴¹. Cells from *Aspm*^{+/+} or *Aspm*^{+/-} and *Aspm*^{-/-} ferrets were merged in a single matrix. An initial run showed that one of the *Aspm*^{+/-} samples contained low Unique Molecular Identifier (UMI) and gene counts compared to the other samples, and clustered differently, so this sample was removed from downstream analyses (Extended Data Fig. 7). Genes were included if they were expressed in ≥ 3 cells and cells were included if they expressed > 200 genes and $< 2,000$ genes. This resulted in 22,211 cells and 21,962 genes from 8,037 *Aspm*^{+/+} or *Aspm*^{+/-} cells and 14,174 *Aspm*^{-/-} cells. The data were log-normalized per cell, scaling each cell to 10,000 molecules as described previously²². The 'MeanVarPlot' Seurat function was used to identify the most variable genes, obtaining 3,555 variable genes. Negative binomial regression was performed on the variable genes, using the number of UMIs per cell as a confounder variable before clustering. The 'PCAfast' function in Seurat was used to implement principal component analysis using the IRLBA package. Twenty-five principal components were selected for clustering and as input for *t*-distributed stochastic neighbour embedding (*t*-SNE) in Seurat. These were selected by plotting the standard deviation of the principal components and setting a cutoff at the 'elbow' of the graph using the 'PCElbowPlot' function in Seurat. Clustering was performed using the Seurat function 'FindClusters', which implements a shared nearest neighbour modularity optimization based algorithm using $k.param = 30$ for defining the k of the k -nearest neighbour algorithm and a resolution of 0.5 as previously described^{41,42}. The Barnes-Hut implementation of *t*-SNE was used for visualizing the clusters using the 'RunTSNE' and 'TSNEPlot' Seurat functions. We observed a co-localization in the *t*-SNE plot of cells clustered together by the graph-based clustering algorithm (Fig. 3a). Cluster markers were obtained with the 'FindAllMarkers' Seurat function using a likelihood-ratio test, using the parameter $min.pct = 0.25$ to test only genes expressed in at least 25% of cells in either all cells or the cells in a specific cluster, and testing only genes with at least 0.25-fold difference on a log-scale between cells in a cluster and all cells using the parameter $thresh.use = 0.25$. *P* values were adjusted for multiple-comparison testing using the $p.adjust$ function in R for the Benjamini-Hochberg false discovery rate (FDR), selecting an FDR threshold of 0.01. Known markers were used to determine the corresponding cell type of each cluster. The heat map in Fig. 3b shows expression data for the top ten cluster markers for each cluster, in a random sample of 10% of the cells of each cluster. Plotting cells onto the *t*-SNE plot based on their batch (three batches with *Aspm*^{+/+} or *Aspm*^{+/-} as well as *Aspm*^{-/-} animals each) suggested that batches did not strongly influence clustering (Extended Data Fig. 7). Plotting cells from each sample onto the *t*-SNE plot suggested that two non-neuronal clusters of blood and choroid plexus epithelial cells were primarily from a single sample (a likely dissection artefact) and these two clusters were removed from further analysis (grey clusters in Fig. 3a and Extended Data Fig. 7c, d). Plotting the number of genes and UMIs in each cluster revealed that one of the excitatory neuronal clusters had almost three times as many genes and UMIs per cell (Extended Data Fig. 7c, d). This, along with the fact that this cluster expressed a combination of markers from the other two excitatory neuronal clusters suggested that this cluster may contain doublets or other technical or batch artefacts; therefore we also removed this cluster from further analysis.

Statistical analysis of cell type composition by genotype. A χ^2 test was performed using the 'chisq.test' function in R to test the association of cluster composition with

genotype. We also quantified the fraction of cells corresponding to each cluster with the assumption that Drop-seq captures and sequences cells in an unbiased manner, and that the frequencies of cells are representative of their frequency in the tissue²². The fraction of cells corresponding to each cluster was obtained by counting the number of cells assigned to each cluster for *Aspm*^{+/+} or *Aspm*^{+/-} and *Aspm*^{-/-} samples and dividing over the total number of cells that passed the filters described above and excluding the three clusters that were removed, for a total of 7,645 *Aspm*^{+/+} and *Aspm*^{+/-} cells and 13,725 *Aspm*^{-/-} cells. Empirical *P* values were obtained by permuting a genotype 10,000 times and obtaining the fraction of cells corresponding to each cluster for each permutation. These fractions were sorted and the *P* value was obtained by counting the number of times a fraction was more extreme or equal to the observed fraction in the non-permuted data divided by 10,000 and multiplied by 2 for a two-tailed test. *P* values were adjusted for multiple comparison testing using the $p.adjust$ function in R for the Benjamini-Hochberg FDR, selecting an FDR threshold of 0.01.

Single-molecule fluorescence in situ hybridization. Using RNAscope fluorescence detection assays and probes (ACDBio), we performed single-molecule fluorescence in situ hybridization according to the manufacturer's protocols. Cryosections on SuperFrost Plus slides (Fisher Scientific) were dried at -20°C , rather than at room temperature or 37°C , for 15 min after mounting, and were used within a week of sectioning. Target retrieval was performed at 80°C in a hybridization oven for 30–40 min before proceeding with the RNAscope multiplex fluorescence detection protocol.

Cell culture and siRNA transfection. H4 and HeLa cells authenticated by short tandem-repeat profiling were obtained from ATCC, cultured in Advanced DMEM (ThermoFisher Scientific) supplemented with 3% FBS (Altantis) and Glutamax-I (ThermoFisher Scientific), and used within five passages with routine mycoplasma screening. Ferret embryonic fibroblasts (FEFs) and mouse embryonic fibroblasts (MEFs) were derived from post-fertilization day 35 and 14.5 embryos, respectively. FEFs and MEFs were cultured in AmnioMAX (ThermoFisher Scientific). H4 cells were transfected with validated siRNAs against human *ASPM* or *PRKCZ*, which encodes PKC ζ (ThermoFisher Scientific), using Oligofectamine and OptiMEM (ThermoFisher Scientific) according to the manufacturer's instructions and were analysed 48 h later.

Immunoprecipitation and immunoblotting. Immunoprecipitation experiments were performed as previously described⁹. In brief, HeLa cells were collected in Dulbecco's PBS (DPBS, ThermoFisher Scientific) and lysed in lysis buffer (50 mM Tris-HCl pH 7.4, 266 mM NaCl, 2.27 mM KCl, 1.25 mM KH_2PO_4 , 6.8 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ and 1% NP-40) supplemented with EDTA-free protease inhibitors cocktail set III (Calbiochem). For each immunoprecipitation, 1 mg of lysate was incubated with 2 μg of antibody for 2 h and then incubated with magnetic protein G-sepharose beads (GE Healthcare Life Sciences) for another 1 h at 4°C . Complexes were washed and then boiled in $2 \times$ Laemmli reducing buffer with β -mercaptoethanol (Bio-Rad). Samples were separated on 4–15% TGX gels (Bio-Rad), transferred onto BA85-supported nitrocellulose (GE Healthcare Life Sciences) at 100 V for 30–45 min using a plate electrode Trans-Blot cell with cooling coil (Bio-Rad) and then subjected to immunoblot analysis using ECL Lightening Plus (Perkin-Elmers) or Western Pico (ThermoFisher Scientific). All immunoprecipitation and immunoblotting experiments were replicated three times.

Fluorescent immunocytochemistry. Cells were fixed in ice-cold methanol for 3 min, permeabilized in blocking buffer (2.5% BSA or FBS, 0.1% Triton X-100, 0.03% NaN_3 in DPBS). Primary and secondary antibodies were diluted in blocking buffer and incubated for 2 h at room temperature. Coverslips were mounted using Gelvatol or Prolong Diamond (ThermoFisher Scientific) and imaged with an inverted confocal microscope (Zeiss LSM700). Images were processed with ImageJ/FIJI. For 3D-structured illumination microscopy (SIM) (Fig. 1e), wild-type and knockout FEFs were plated on 1.5-mm coverslips and immunostained as above. Coverslips were mounted with Vectashield (Vectorlabs). 3D-SIM imaging was performed on a Zeiss Elyra PS.1 microscope equipped with a 100×1.40 NA oil objective. Exciting light was directed through a movable optical grating to generate a fine-striped interference pattern on the same plane. *z* stacks of 15 optical sections with a step size of $0.1 \mu\text{m}$ were acquired to generate images in maximum intensity projection. The epitope of the ASPM (216-1) antibody⁴³, NDNYGLNQDLESES, is located before the TALEN target site. The following antibodies were used at 1:100–1:2,000: centrin (Millipore 20H5), PAR6 α (SCBT sc-14405), PAR6 α (Abcam ab180159), β -actin (Proteintech 20536-1-AP), ASPM (SCBT sc-98903), ASPM (gift from J. Bond, 216-1), ninein (Biolegend Poly6028) and aPKC ζ (SCBT sc-216).

Statistics and reproducibility. All experiments in Fig. 1e, 1, m, 2a–i, 1–q, 3d, 4a–g were repeated independently three times with similar results. No statistical method was used to predetermine sample size. At least three animals or samples were generally analysed per genotype or age. Two-tailed *t*-tests were performed for most data using Prism 7, unless otherwise stated. Ferret kits were born at a Mendelian ratio but the genotype of each individual kit was random, which inherently randomized our experiments. To perform blinded experiments, the genotype

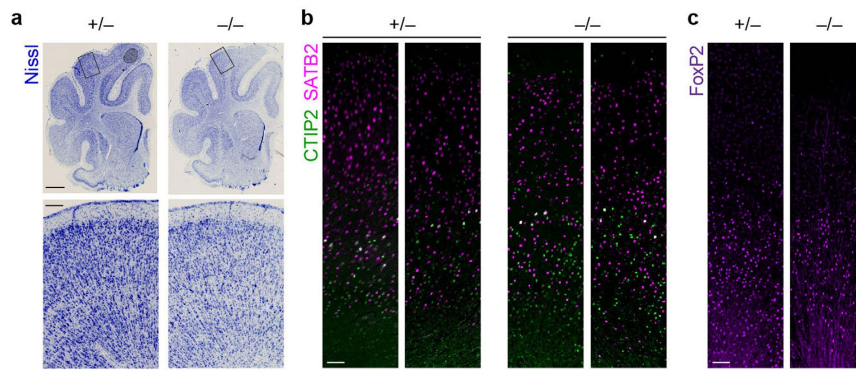
of each animal was revealed only after the analysis was completed. In ferrets, sex was undifferentiated up to P21, after which only male ferrets were analysed. In mice, sex was undifferentiated up to P0, after which only male mice were analysed. Figure 1j, k, $n = 3$ male ferrets per genotype of > 8 months of age. Individual P values can be found in the Source Data associated with this figure. Figure 1n, box plot elements are maximum, third quartile, median, first quartile and minimum. E35: $n = 17$ *Aspm*^{+/+} or *Aspm*^{+/-} ferrets and $n = 10$ *Aspm*^{-/-} ferrets from three litters (significant differences in brain weight between *Aspm*^{+/+} or *Aspm*^{+/-} and *Aspm*^{-/-} ferrets are indicated: $P = 0.0023$); P0: $n = 9$ *Aspm*^{+/+} or *Aspm*^{+/-} ferrets and $n = 6$ *Aspm*^{-/-} ferrets from one litter ($P = 0.0003$); P21/22 (3 weeks): $n = 8$ *Aspm*^{+/+} or *Aspm*^{+/-} ferrets and $n = 3$ *Aspm*^{-/-} ferrets from two litters ($P = 0.0010$); P41 and older animals (> 6 weeks): $n = 7$ *Aspm*^{+/+} or *Aspm*^{+/-} ferrets and $n = 7$ *Aspm*^{-/-} ferrets ($P = 0.0094$). Because brain weight was not found to be significantly different after P41, both adult and P41 animals were combined into a ' > 6 weeks' group. Ferrets display considerable variability in body weight and brain weight at birth, related to variance in the exact time of birth post-conception and to litter size, which can vary from 3 to 15 kits. Thus, one small litter of three P0 kits, including one wild-type and two *Aspm*^{+/-} young, which had body weights $\sim 2\times$ of the other P0 litters collected, were excluded from brain weight analysis. Because the y axis is log scale, overlaying each data point as dot plots for $n < 10$ does not indicate the distribution of the data efficiently. Instead, brain weight of individual animals can be found in the Source Data associated with this figure. Figure 1o, using the whole-brain images of coronal sections stained with Nissl or DAPI from $n = 6$ animals per genotype, we manually measured mean cortical thickness of the posterior sigmoid gyrus. No significant difference was found ($P = 0.0843$). Figure 1p, the same animals used for MRI (3 *Aspm*^{+/-} and 3 *Aspm*^{-/-} ferrets as described above) were used for body weight analysis. No significant difference was found ($P = 0.4481$). Figure 2j, immunofluorescence images were coded and counted blind to genotype by four individuals, and the four independent counts were then averaged for each brain section. The inter-individual correlation was $r \geq 0.89$. Four to six brain sections were imaged and counted per animal, and $n = 3$ *Aspm*^{+/+} or *Aspm*^{+/-} and $n = 4$ *Aspm*^{-/-} littermate E35 animals were analysed. Figure 3f, *Apod*⁺ cells from single-molecule fluorescence in situ hybridization were segmented and counted using ImageJ, in an area of dorsal cortex $400 \times 400 \mu\text{m}^2$ centred on the intermediate zone, in multiple sections per animal, with $n = 4$ *Aspm*^{+/+} or *Aspm*^{+/-} ferrets and $n = 4$ *Aspm*^{-/-} ferrets at E35. Per brain average counts were then compared using a one-tailed t -test.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The code used in this study is available from the corresponding author upon reasonable request.

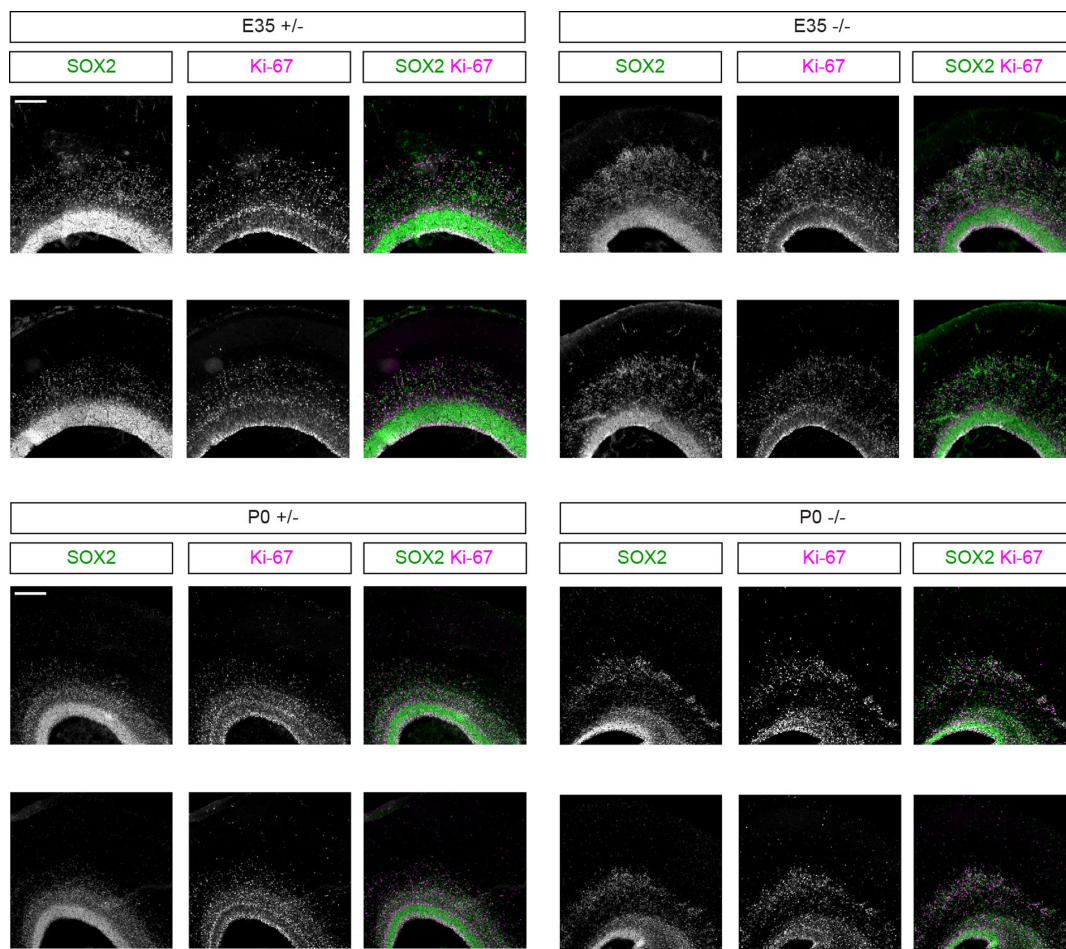
Data availability. scRNA-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE110010. All other data are included in the paper (Source Data for Figs. 1–3) and in the Supplementary Information.

30. Mashiko, D. et al. Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep.* **3**, 3355 (2013).
31. Li, Z., Sun, X., Chen, J., Leno, G. H. & Engelhardt, J. F. Factors affecting the efficiency of embryo transfer in the domestic ferret (*Mustela putorius furo*). *Theriogenology* **66**, 183–190 (2006).
32. Chahboune, H. et al. Neurodevelopment of C57B/L6 mouse brain assessed by in vivo diffusion tensor imaging. *NMR Biomed.* **20**, 375–382 (2007).
33. Petersen, K. F. et al. The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc. Natl Acad. Sci. USA* **104**, 12587–12594 (2007).
34. Rueckert, D. et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).
35. Papademetris, X., Jackowski, A. P., Schultz, R. T., Staib, L. H. & Duncan, J. S. Integrated intensity and point-feature nonrigid registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* Vol. 3216 (eds Barillot, C., et al.) 763–770 (Springer, Berlin, 2001).
36. Sawada, K. & Watanabe, M. Development of cerebral sulci and gyri in ferrets (*Mustela putorius*). *Congenit. Anom. (Kyoto)* **52**, 168–175 (2012).
37. Joshi, A. et al. Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* **9**, 69–84 (2011).
38. Rahman, A. S., Parvinjah, S., Hanna, M. A., Helguera, P. R. & Busciglio, J. Cryopreservation of cortical tissue blocks for the generation of highly enriched neuronal cultures. *J. Vis. Exp.* **45**, e2384 (2010).
39. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
40. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
41. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
42. Villani, A. C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
43. Higgins, J. et al. Human ASPM participates in spindle organisation, spindle orientation and cytokinesis. *BMC Cell Biol.* **11**, 85 (2010).



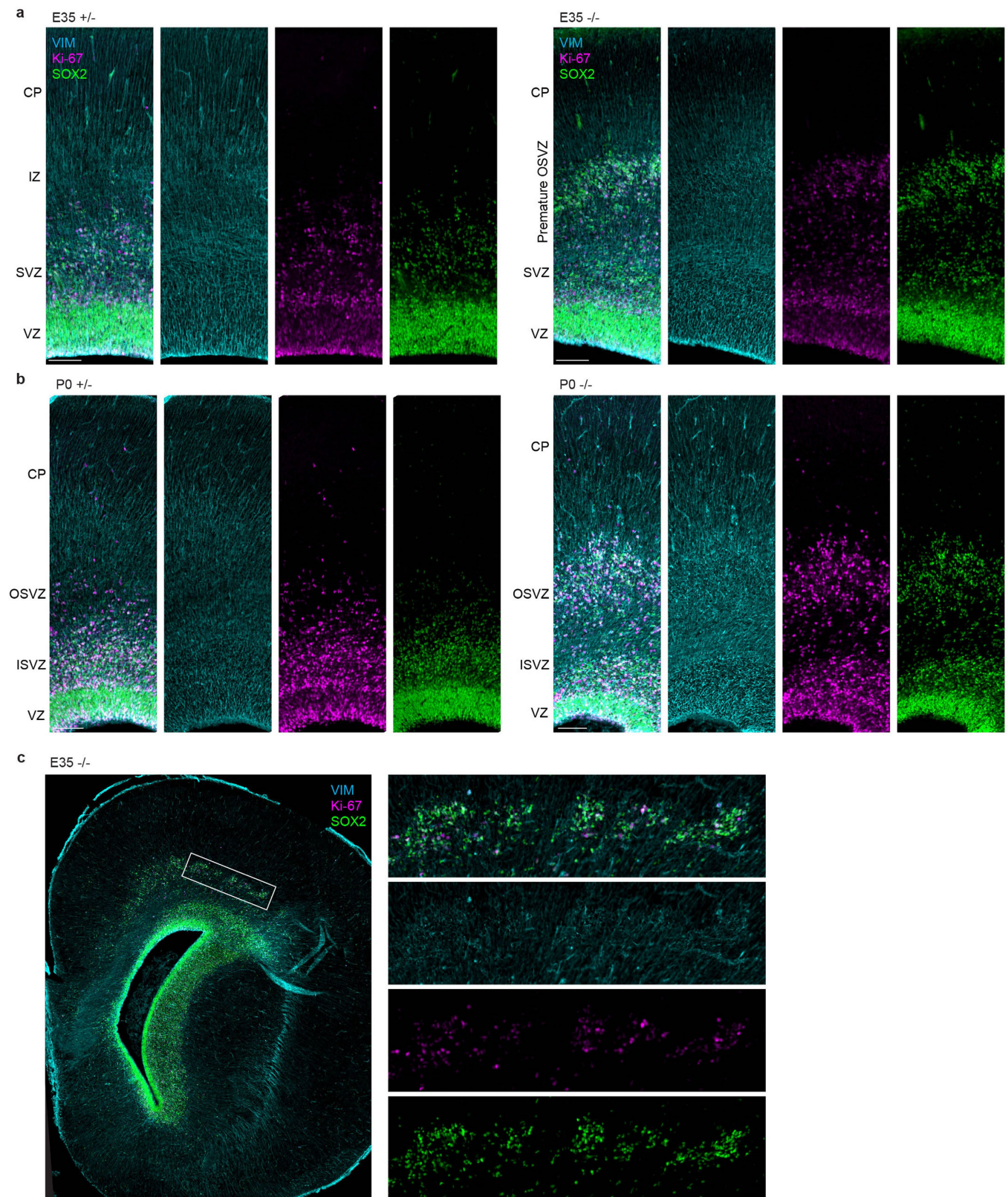
Extended Data Fig. 1 | Cytoarchitecture and neuronal subtype lamination in the cortex of mature *Aspm* knockout ferrets. a, Nissl stains of coronal sections from the brains of P41 littermates, as shown in Fig. 1l, with additional *Aspm*^{+/−} and *Aspm*^{−/−} littermates shown. **b, c**, Brain sections of P41 littermates immunostained for cortical layer-specific

projection neurons including SATB2 (layer II–IV), CTIP2 (layer V), and FoxP2 (layer VI). The experiments were repeated independently three times with similar results. Scale bars, 2 mm (**a**, top), 200 μm (**a**, bottom) and 100 μm (**b**, **c**).



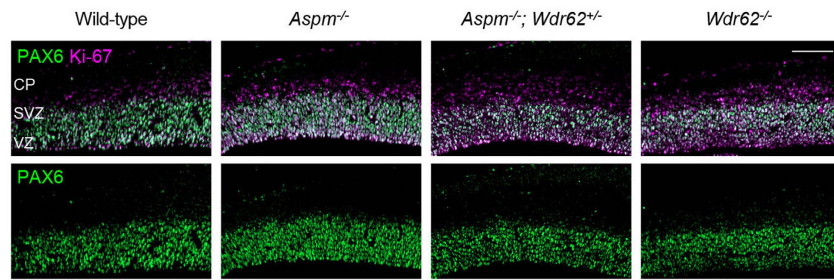
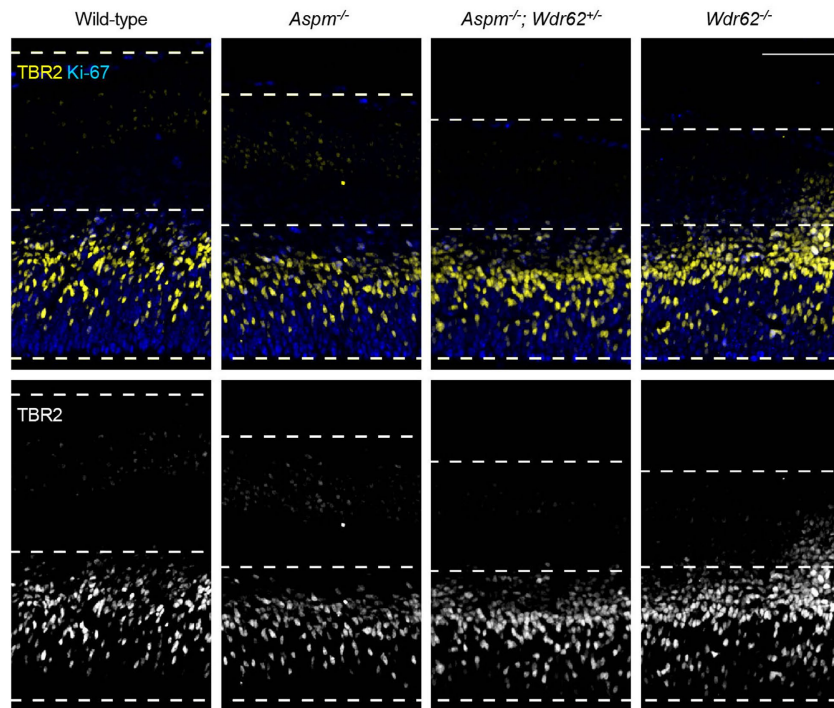
Extended Data Fig. 2 | SOX2 and Ki-67 immunostaining in additional E35 and P0 littermates dorsal cortex. Additional results to Fig. 2e, f. Each set of panels is from the brain of a different littermate, showing the high

penetrance of the neural progenitor cell basal displacement phenotype. The experiments were repeated independently three times with similar results. Scale bar, 200 μ m.



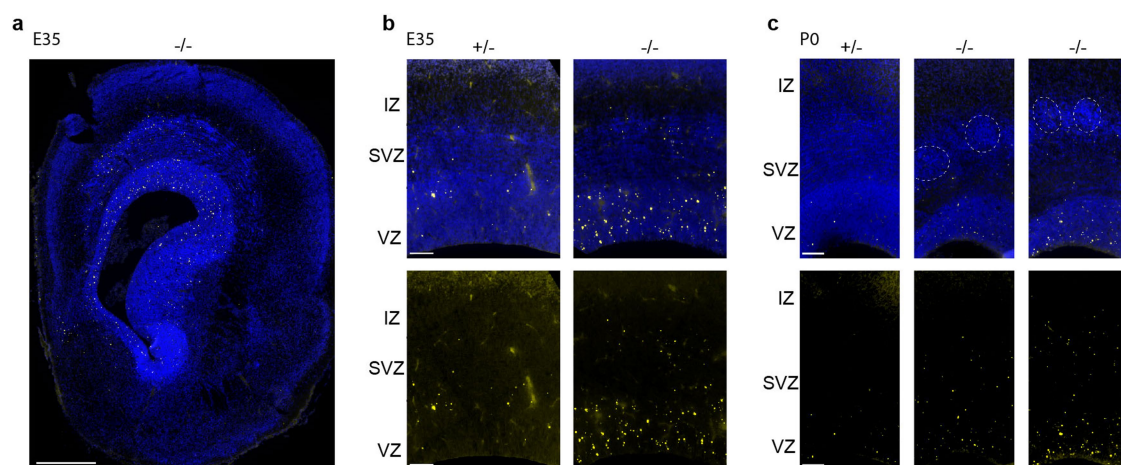
Extended Data Fig. 3 | Displaced progenitors in *Aspm* knockout ferrets have basal fibres. Additional results for Fig. 2h, i. Immunostaining of SOX2, Ki-67 and VIM shows that displaced neural progenitors have basal

radial fibres. The experiments were repeated independently three times with similar results. Scale bars, 100 μ m.

a E14.5 mouse**b** E14.5 mouse

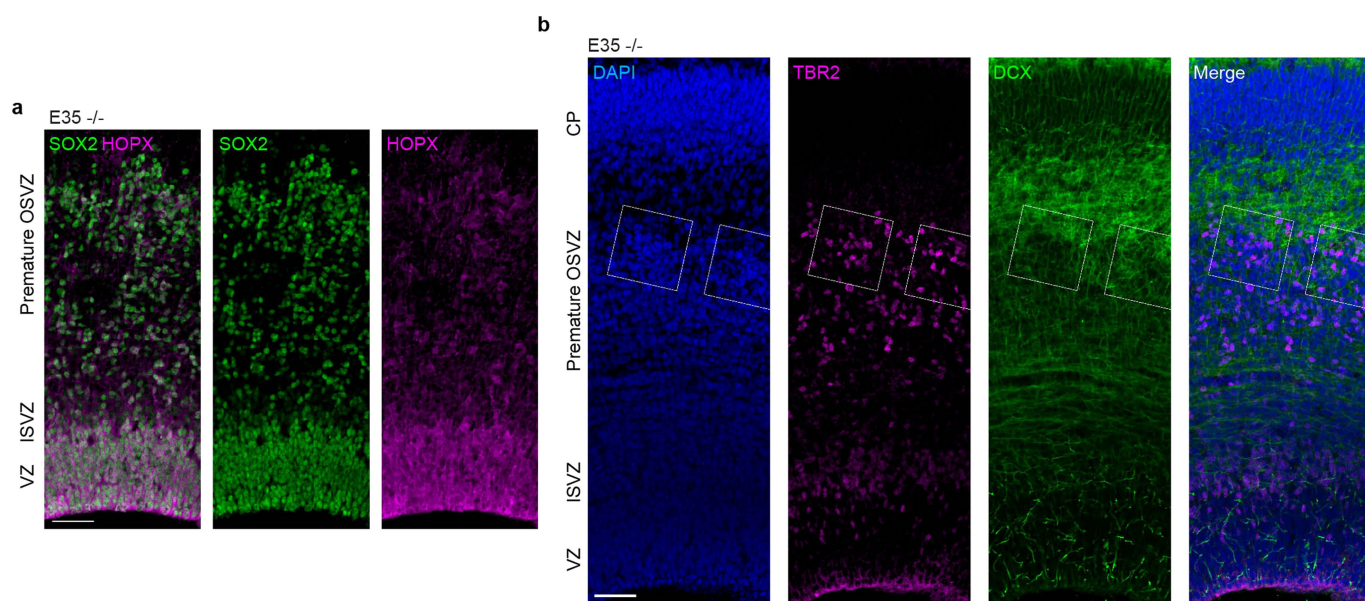
Extended Data Fig. 4 | *Aspm* knockout mice do not demonstrate displaced progenitors in the intermediate zone. a, b, Unlike *Aspm*^{-/-} ferrets, *Aspm*^{-/-} mice do not have displaced NPCs in the intermediate zone. However, they show a variable increase in the number of intermediate progenitors (PAX6⁻Ki-67⁺ cells in **a** and TBR2⁺ cells in **b**),

which is enhanced by heterozygous, compound mutation in *Wdr62*, a microcephaly gene causing more severe microcephaly⁹. The experiments were repeated independently three times with similar results. Scale bars, 100 μ m.



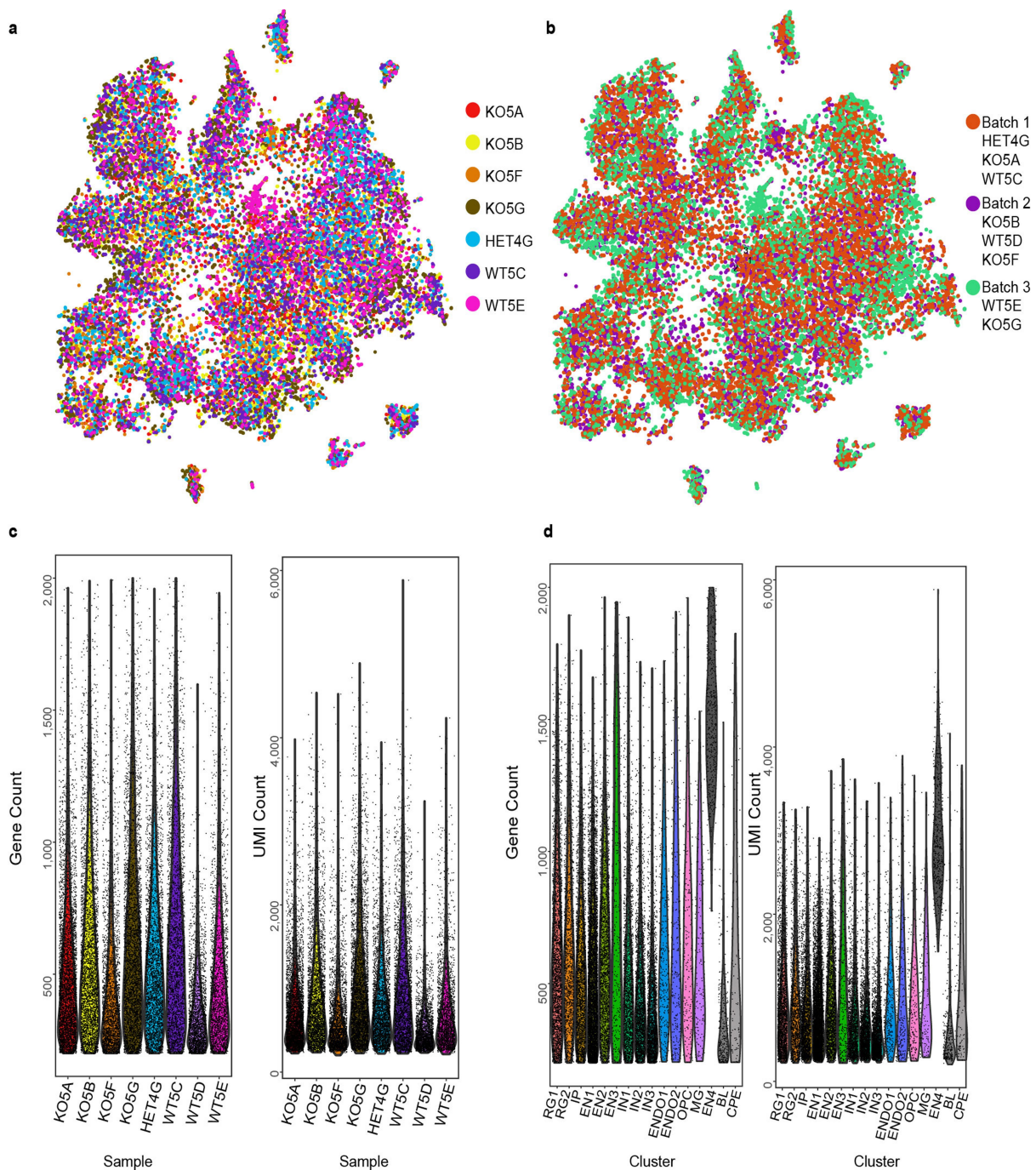
Extended Data Fig. 5 | Modest increase in apoptosis throughout the germinal zones of the *Aspm* knockout telencephalon. Apoptotic cells (yellow) are indicated by enzymatic fluorescence detection of double-stranded DNA damage with DAPI nuclear counterstaining (blue).

The experiments were repeated independently three times with similar results. **a**, Whole section. **b**, **c**, Cortical wall columns. Scale bars, 500 μm (**a**) and 100 μm (**b**, **c**).



Extended Data Fig. 6 | Additional immunohistochemical analyses of displaced progenitors in the *Aspm* knockout cortex. **a**, E35 knockout cortex stained for VRG and ORG markers SOX2 and HOPX reveals extensive co-labelling in both the ventricular zone (VZ) and SVZ, including in displaced OSVZ progenitors. **b**, In the E35 knockout OSVZ, clusters of supernumerary displaced neural progenitor cells include

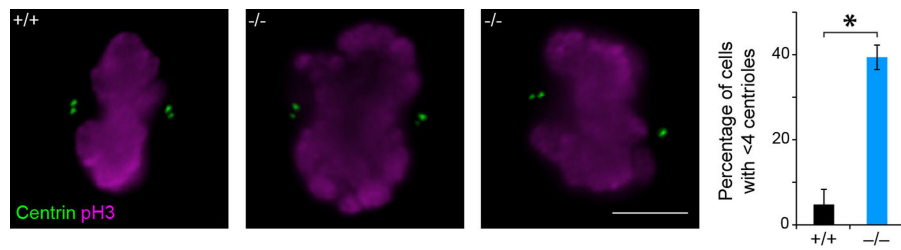
numerous TBR2⁺ intermediate progenitors and are surrounded by DCX⁺ newborn neurons, indicating preserved neurogenesis within the precocious OSVZ niche of the *Aspm* knockout cortex. The experiments were repeated independently three times with similar results. Scale bars, 50 μ m.



Extended Data Fig. 7 | scRNA-seq batch, sample and cluster analyses.

a, *t*-SNE plot from Fig. 3a with cells coloured by biological replicate (that is, animal). Most clusters include cells from all samples, except for a cluster expressing blood genes and a cluster expressing choroid plexus epithelial cells that are mostly from animal WT5E. These two cell clusters were not included in downstream analyses. HET, heterozygote; KO, knockout; WT, wild type. Numbers and letters indicate litter and animal identification number, respectively. **b**, *t*-SNE plot from Fig. 3a with cells coloured by the batch they were processed in. Clusters are composed of cells from all batches. **c**, Per-cell gene count and UMI count per sample. Each violin plot is one biological replicate and each dot is one cell. Sample WT5D was not included in the analysis due to the lower gene and UMI count compared to other samples as well as the inconsistent clustering compared to other wild-type samples (data not shown). **d**, Per-cell gene count and

UMI count for identified clusters. Each violin plot is one cell cluster and each dot is one cell. The three clusters in grey (EN4, BL, CPE) were not included in downstream analyses. See Methods for details. This scRNA-seq experiment was performed once with $n = 22,211$ cells (8,037 cells from two *Aspm*^{+/+} and one *Aspm*^{-/-} ferrets and 14,174 cells from four *Aspm*^{-/-} ferrets). RG1, cycling radial glial progenitors; RG2, interphase radial glial progenitors; IP, intermediate progenitors; EN1, upper-layer excitatory neurons; EN2, deep-layer excitatory neurons; EN3, Cajal–Retzius cells; IN1, immature inhibitory neurons; IN2, SST⁺ inhibitory neurons; IN3, ventral/inhibitory progenitors; ENDO1, endothelial cells 1; ENDO2, endothelial cells 2; OPC, oligodendrocyte precursors; MG, microglia; EN4, mixed excitatory neuron identity; BL, blood cells; CPE, choroid plexus epithelial cells.



Extended Data Fig. 8 | Loss of *Aspm* disrupts centriole duplication in FEFs. Mitotic *Aspm* knockout FEFs, identified by staining for pH3 and co-stained for the centriolar marker centrin, display a significant loss of centrioles. The percentage of cells with an abnormal number (less than 4) of centrioles is increased eightfold in *Aspm*^{-/-} FEFs compared to *Aspm*^{+/+}

FEFs ($n = 100$ cells per genotype for three independent experiments; $P = 0.003$). The experiments were repeated independently three times with similar results. Statistical analysis was performed using a two-tailed t -test; data are mean \pm s.e.m.

Extended Data Table 1 | Region-specific changes in volume and surface area by loss of *Aspm* in ferrets

| a Volume (mm ³) | | | | b Outer surface area (mm ²) | | | |
|------------------------------------|----------------------------|----------------------------|-----------------|--|----------------------------|----------------------------|-----------------|
| Region | <i>Aspm</i> ^{+/−} | <i>Aspm</i> ^{−/−} | <i>P</i> -value | Region | <i>Aspm</i> ^{+/−} | <i>Aspm</i> ^{−/−} | <i>P</i> -value |
| Frontal ctx | 475.0 ± 58.4 | 235.0 ± 11.0 | 0.0156 | Cerebral ctx (total) | 1,225.4 ± 64.1 | 933.0 ± 47.5 | 0.0217 |
| Corpus callosum | 79.3 ± 6.7 | 40.3 ± 2.3 | 0.0048 | Frontal ctx | 309.6 ± 18.5 | 203.5 ± 2.8 | 0.0048 |
| Lateral ctx | 666.9 ± 48.4 | 364.8 ± 33.9 | 0.0069 | Lateral ctx | 391.9 ± 18.5 | 301.4 ± 20.5 | 0.0305 |
| Ctx WM | 701.4 ± 46.1 | 402.9 ± 19.0 | 0.0039 | Parietal/occipital ctx | 298.4 ± 20.1 | 217.1 ± 22.7 | 0.0551 (NS) |
| Parietal/occipital ctx | 421.6 ± 17.5 | 256.6 ± 17.5 | 0.0026 | Medial ctx | 225.5 ± 9.7 | 211.0 ± 9.5 | 0.3476 (NS) |
| Medial ctx | 639.4 ± 58.5 | 397.3 ± 14.3 | 0.0159 | | | | |
| Hippocampus | 319.7 ± 49.8 | 208.4 ± 18.5 | 0.0136 | | | | |
| Caudate | 113.9 ± 13.3 | 74.8 ± 4.5 | 0.0478 | | | | |
| Putamen | 14.3 ± 0.9 | 10.1 ± 0.4 | 0.0147 | | | | |
| Thalamus | 239.7 ± 18.5 | 172.4 ± 3.4 | 0.0238 | | | | |
| Cerebellum GM | 895.0 ± 32.0 | 650.3 ± 20.7 | 0.0030 | | | | |
| Cerebellum WM | 182.5 ± 12.1 | 141.5 ± 6.2 | 0.0382 | | | | |
| Brainstem GM | 263.8 ± 23.1 | 207.0 ± 8.4 | 0.0836 (NS) | | | | |
| Brainstem WM | 149.8 ± 6.5 | 121.3 ± 5.7 | 0.0298 | | | | |
| Midbrain WM | 43.2 ± 0.6 | 37.3 ± 0.8 | 0.0018 | | | | |
| Amygdala | 23.2 ± 1.1 | 20.2 ± 1.4 | 0.1868 (NS) | | | | |
| Midbrain GM | 106.0 ± 3.2 | 101.8 ± 2.6 | 0.3146 (NS) | | | | |

| d Fractional anisotropy | | | |
|--------------------------------|----------------------------|----------------------------|-----------------|
| Region | <i>Aspm</i> ^{+/−} | <i>Aspm</i> ^{−/−} | <i>P</i> -value |
| Frontal ctx | 0.272 ± 0.002 | 0.240 ± 0.007 | 0.0130 |
| Medial ctx | 0.255 ± 0.004 | 0.237 ± 0.009 | 0.1377 (NS) |
| Lateral ctx | 0.253 ± 0.011 | 0.252 ± 0.008 | 0.9198 (NS) |
| Parietal/occipital ctx | 0.267 ± 0.014 | 0.246 ± 0.014 | 0.3421 (NS) |
| Ctx WM | 0.431 ± 0.031 | 0.379 ± 0.018 | 0.2285 (NS) |
| Corpus callosum | 0.487 ± 0.034 | 0.434 ± 0.048 | 0.4170 (NS) |
| Cingulum | 0.415 ± 0.063 | 0.342 ± 0.053 | 0.4267 (NS) |
| Fornix | 0.458 ± 0.041 | 0.435 ± 0.022 | 0.6542 (NS) |
| Anterior commissure | 0.437 ± 0.069 | 0.336 ± 0.052 | 0.3046 (NS) |
| Optic tract | 0.523 ± 0.059 | 0.435 ± 0.045 | 0.3013 (NS) |
| Corticospinal tract | 0.550 ± 0.059 | 0.470 ± 0.041 | 0.3330 (NS) |
| Brainstem WM | 0.501 ± 0.042 | 0.423 ± 0.026 | 0.1925 (NS) |
| Cerebellum WM | 0.454 ± 0.031 | 0.368 ± 0.033 | 0.1321 (NS) |
| Midbrain WM | 0.398 ± 0.058 | 0.338 ± 0.043 | 0.4496 (NS) |
| Cerebellum GM | 0.301 ± 0.010 | 0.258 ± 0.007 | 0.0211 |
| Brainstem GM | 0.353 ± 0.011 | 0.285 ± 0.015 | 0.0230 |
| Inferior colliculus | 0.289 ± 0.001 | 0.276 ± 0.009 | 0.2078 (NS) |
| Superior colliculus | 0.270 ± 0.008 | 0.262 ± 0.008 | 0.4870 (NS) |
| Periaqueductal gray | 0.273 ± 0.010 | 0.268 ± 0.003 | 0.6517 (NS) |
| Midbrain GM | 0.291 ± 0.017 | 0.270 ± 0.015 | 0.3974 (NS) |
| Thalamus | 0.286 ± 0.015 | 0.269 ± 0.003 | 0.3413 (NS) |
| Hippocampus | 0.270 ± 0.011 | 0.254 ± 0.002 | 0.2276 (NS) |
| Hypothalamus | 0.257 ± 0.021 | 0.225 ± 0.011 | 0.2533 (NS) |
| Globus pallidus | 0.291 ± 0.020 | 0.288 ± 0.021 | 0.9115 (NS) |
| Putamen | 0.285 ± 0.007 | 0.292 ± 0.023 | 0.7631 (NS) |
| Caudate | 0.216 ± 0.005 | 0.221 ± 0.013 | 0.7161 (NS) |
| Septum | 0.267 ± 0.022 | 0.245 ± 0.010 | 0.4183 (NS) |
| Amygdala | 0.206 ± 0.013 | 0.213 ± 0.019 | 0.7790 (NS) |

a, Multiple brain regions are significantly decreased in volume; the highest reduction was found in the frontal cortex of adult *Aspm*^{−/−} ferrets (*n* = 3 per genotype). Subcortical regions were relatively preserved. **b**, The outer cortical surface is the most reduced in the frontal cortex followed by the lateral cortex. The parietal/occipital cortex is also decreased but the difference was not significant. The medial cortex shows no discernible decrease. **c**, **d**, DTI shows that the orientation of white matter tracts or connectivity is fundamentally unchanged in *Aspm* knockout ferrets except in the frontal cortex, which shows a modest decrease in fractional anisotropy (**d**). The directional map (**c**) shows white matter orientation. Red, green and blue indicate the medial–lateral, superior–inferior, and anterior–posterior components, respectively. Statistical analysis was performed using a two-tailed *t*-test. Data are mean ± s.e.m. NS, not significant.

Extended Data Table 2 | Cluster identifiers of E35 ferret cerebral cortical cells analysed by scRNA-seq

| Color | Cluster ID | Top 10 Marker Transcripts | Inferred Cell Type | Cell Count in <i>Aspm</i> ^{+/+} | Cell Count in <i>Aspm</i> ^{-/-} | Proportion of <i>Aspm</i> ^{+/+} | Proportion of <i>Aspm</i> ^{-/-} | Fold-change in proportion | Empirical P-value |
|-------|------------|---|-------------------------------------|--|--|--|--|---------------------------|-------------------|
| | RG1 | TOP2A; HMGB2; CENPF; CENPE; TPX2; XLOC_000183; 2810417H13Rik; SMC4; XLOC_036181; KIF11 | Cycling radial glial progenitors | 803 | 1,661 | 10.5% | 12.1% | +15% | 0.0002 (FDR<0.01) |
| | RG2 | VIM; HES1; SLC1A3; GDDP2; SFRP1; PON2; PTN; SMPDL3A; NES; PAX6 | Interphase radial glial progenitors | 495 | 800 | 6.5% | 5.8% | -10% | 0.0696 |
| | IP | NRN1; PTPDC1; TENM4; ELAVL4; ROBO2; FBXW7; IGFBP2; NELL2; PRKX; TTC28 | Intermediate progenitors | 686 | 1,603 | 9.0% | 11.7% | +30% | 0.0002 (FDR<0.01) |
| | EN1 | SYT4; CSRP2; NEUROD6; NEUROD2; UNC5D; NTM; NSG2; LIMCH1; SORBS2; ISLR2 | Upper-layer excitatory neurons | 2,203 | 3,490 | 28.8% | 25.4% | -12% | 0.0002 (FDR<0.01) |
| | EN2 | NEFM; FEZF2; NEFL; GRIA2; GUCY1B3; DYNC11I; ARPP21; NEUROD6; B3GALT2; KCTD12 | Deep-layer excitatory neurons | 981 | 1,053 | 12.8% | 7.7% | -40% | 0.0002 (FDR<0.01) |
| | EN3 | RELN; LHX1; NDNF; TP73; NHLH2; VSNL1; ENSMPUG00000012124; PLCL1; ENSMPUG00000009767; SEMA6A | Cajal-Retzius cells | 103 | 227 | 1.3% | 1.7% | +23% | 0.0704 |
| | IN1 | PBX3; MEIS2; XLOC_008478; ENSMPUG00000024751; INA; MAP1B; ATP1B1; XLOC_021005; RUNX1T1; JAKMIP2 | Immature inhibitory neurons | 591 | 1,167 | 7.7% | 8.5% | +10% | 0.0444 |
| | IN2 | SST; XLOC_014564; PDZRN4; NXPH1; NXPH2; SYT1; XLOC_026835; GRIA1; ARX; PDE4DIP | SST interneurons | 734 | 1,568 | 9.6% | 11.4% | +19% | 0.0002 (FDR<0.01) |
| | IN3 | XLOC_026835; DLX1; XLOC_026893; CCDC88A; ZNF704; PFN2; EPHA5; XLOC_007250; INA; NR2F1 | Ventral/inhibitory progenitors | 710 | 1,521 | 9.3% | 11.1% | +19% | 0.0002 (FDR<0.01) |
| | ENDO1 | APOA1; COL4A1; CALD1; SPARC; FN1; IGFBP7; XLOC_010971; ENSMPUG00000012145; LAMA4; MGP | Endothelial cells | 122 | 210 | 1.6% | 1.5% | -4% | 0.7418 |
| | ENDO2 | SPARCL1; LYZ; PECAM1; IFNAR1; SPARC; IFI27; XLOC_010971; EMB; S100A6; IGFBP7 | Endothelial cells | 113 | 150 | 1.5% | 1.1% | -26% | 0.0166 (FDR<0.05) |
| | OPC | APOD; XLOC_017682; SCRG1; DBI; SPARCL1; PDGFRA; SERPINE2; ENSMPUG00000011077; PTPRZ1; LHFPL3 | Oligodendrocyte precursors | 48 | 170 | 0.6% | 1.2% | +97% | 0.0002 (FDR<0.01) |
| | MG | ENSMPUG00000001122; C1QC; C3; RGS10; XLOC_039347; CCL8; XLOC_039690; ZFP36; CCL4; SPP1 | Microglia | 56 | 105 | 0.7% | 0.8% | +4% | 0.7234 |

The three clusters highlighted in blue represent the largest proportional changes with empirical FDR < 0.01, and are similarly indicated in Fig. 3. Statistical analysis was performed using a two-tailed χ^2 test.

An evolutionarily conserved ribosome–rescue pathway maintains epidermal homeostasis

Kifayathullah Liakath^{1,8}, Eric W. Mills², Inês Sequeira¹, Beate M. Lichtenberger^{1,3}, Angela Oliveira Pisco¹, Kalle H. Sipilä¹, Ajay Mishra^{1,4}, Harunori Yoshikawa⁵, Colin Chih-Chien Wu², Tony Ly^{5,6}, Angus I. Lamond⁵, Ibrahim M. Adham⁷, Rachel Green² & Fiona M. Watt^{1*}

Ribosome-associated mRNA quality control mechanisms ensure the fidelity of protein translation^{1,2}. Although these mechanisms have been extensively studied in yeast, little is known about their role in mammalian tissues, despite emerging evidence that stem cell fate is controlled by translational mechanisms^{3,4}. One evolutionarily conserved component of the quality control machinery, Dom34 (in higher eukaryotes known as Pelota (Pelo)), rescues stalled ribosomes⁵. Here we show that Pelo is required for mammalian epidermal homeostasis. Conditional deletion of *Pelo* in mouse epidermal stem cells that express Lrig1 results in hyperproliferation and abnormal differentiation of these cells. By contrast, deletion of *Pelo* in Lgr5-expressing stem cells has no effect and deletion in Lgr6-expressing stem cells induces only a mild phenotype. Loss of *Pelo* results in accumulation of short ribosome footprints and global upregulation of translation, rather than affecting the expression of specific genes. Translational inhibition by rapamycin-mediated downregulation of mTOR (mechanistic target of rapamycin kinase) rescues the epidermal phenotype. Our study reveals that the ribosome-rescue machinery is important for mammalian tissue homeostasis and that it has specific effects on different stem cell populations.

Pelo is expressed in mouse skin dermis and epidermis⁶ (Extended Data Fig. 1a). Dermal-specific deletion of *Pelo* (*Pelo*^{derKO}) resulted in mice that were smaller than littermate controls but had a normal lifespan and no dermal abnormalities (Fig. 1a–f). Although *Pelo* forms a functional complex with Hbs1 in yeast⁷ and the mammalian homologue Hbs1l is expressed in mouse skin⁶ (Extended Data Fig. 1b), *Hbs1l* knockout (from exon 5; Extended Data Fig. 1c) caused no epidermal defects (Extended Data Fig. 1d–f) and only small changes in dermal collagen deposition, thickness and cell density (Extended Data Fig. 1f–m). Another *Pelo* partner, Gtpbp2⁸, does not have a reported skin phenotype.

Selective embryonic deletion of *Pelo* in Krt14-expressing epidermal cells, comprising the known stem cell subpopulations⁹, via *Krt14*^{Cre} (*Pelo*^{epiKO}; Fig. 1g) phenocopied deletion via the ubiquitous *Rosa26* locus¹⁰. Mice were born with scaly skin and an epidermal barrier defect (increased transepidermal water loss (TEWL)). They exhibited hair and weight loss and failed to thrive beyond 5 months of age (Fig. 1h–k). Epidermal thickening resulted from increased proliferation (Fig. 1l–s) and abnormal accumulation of differentiated cells (Fig. 1n–t). Wound closure was delayed (Fig. 1u), correlating with reduced proliferation, differentiation and migration of epidermal cells (Extended Data Fig. 2a–i). Hyperproliferation in unwounded skin combined with delayed wound healing and abnormal differentiation has been observed in other mouse models¹¹. There was also striking degeneration of the sebaceous glands and hair follicles, correlating with loss of the hair follicle bulge stem cell markers Krt15

and CD34 and the junctional zone stem cell marker Lrig1 (Extended Data Fig. 3a–c).

To determine whether the *Pelo* epidermal phenotype could be induced postnatally, we induced epidermal loss of *Pelo* in adult mice by treating adult *Pelo*^{fl/fl}*Krt14*^{CreERT} mice with 4-hydroxytamoxifen (4-OHT; Extended Data Fig. 4a, b). Mice developed skin lesions, increased TEWL and delayed wound closure (Extended Data Fig. 4c–e). Degeneration of hair follicles and sebaceous glands correlated with keratinized cyst formation (Extended Data Fig. 4f, g). Sebocyte differentiation was disturbed, accompanied by expansion of Lrig1 labelling into the upper sebaceous gland (Extended Data Fig. 4h, i).

PELO knockdown in cultured human epidermal keratinocytes led to an increase in stem cell colonies (Extended Data Fig. 5a–g). Immunostaining of epidermis reconstituted on decellularized dermis revealed increased proliferation of basal layer cells and increased differentiated layers (Extended Data Fig. 5h–l). Therefore, the mouse epidermal *Pelo* phenotype was recapitulated in human cells.

To determine whether there is a differential requirement for *Pelo* in different epidermal subpopulations, we conditionally deleted *Pelo* in Lgr5⁺, Lgr6⁺ and Lrig1⁺ stem cells (Fig. 2a–c). *Pelo* deletion in Lrig1⁺ cells recapitulated the effects of deleting *Pelo* in Krt14⁺ cells, whereas when *Pelo* was deleted in Lgr5⁺ and Lgr6⁺ cells differentiation was normal (Fig. 2d) with only a small increase in Ki67⁺ cells (Fig. 2f, Extended Data Fig. 5m). *Pelo* deletion in Lrig1⁺ cells increased cell proliferation in the upper hair follicle, with marked changes in follicles and sebaceous glands (Fig. 2e, Extended Data Fig. 6a, b). There were substantial increases in proliferation and TEWL in the interfollicular epidermis (IFE) of *Pelo*^{fl/fl}*Lrig1*^{CreERT2} mice compared to *Pelo*^{fl/fl}*Lgr5*^{CreERT2} and *Pelo*^{fl/fl}*Lgr6*^{CreERT2} mice (Extended Data Fig. 5m, Fig. 2f, h). There was a small increase in epidermal thickness in *Pelo*^{fl/fl}*Lgr6*^{CreERT2} mice but TEWL was unaffected (Fig. 2g, h).

We next generated *Pelo*^{fl/fl}*Lrig1*^{CreERT2}*Rosa26*^{tdTom}, *Pelo*^{fl/fl}*Lgr5*^{CreERT2}*Rosa26*^{tdTom} and *Pelo*^{fl/fl}*Lgr6*^{CreERT2}*Rosa26*^{tdTom} mice, and treated them with 4-OHT. *Pelo* deletion did not change the contribution of Lgr5 or Lgr6 progeny to the epidermis (Extended Data Fig. 6c, d). By contrast, on *Pelo* deletion Lrig1 lineage cells expanded downwards into the hair follicles and fully colonized the IFE (Extended Data Fig. 6c, d). In the presence or absence of *Pelo*, the Lrig1 lineage accounted for most Ki67⁺ epidermal cells; they also accounted for the increase in proliferative cells on *Pelo* deletion (Extended Data Fig. 6e, f).

Yeast cells lacking *Dom34* (the homologue of *Pelo*) are enriched in short 16–18-nucleotide ribosome-protected fragments (RPFs) resulting from translation to the 3' end of truncated mRNAs⁵. *Dom34* and *Rli1* mutant yeast cells accumulate full length 28–32-nucleotide RPFs in 3' untranslated regions (UTRs), consistent with the roles of *Dom34* and *Rli1* in ribosome rescue and recycling on intact mRNAs, respectively¹². In nucleate haematopoietic cells, PELO and ABCE1 (mammalian

¹Centre for Stem Cells and Regenerative Medicine, King's College London, London, UK. ²Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, MD, USA. ³Skin and Endothelium Research Division, Department of Dermatology, Medical University of Vienna, Vienna, Austria. ⁴Cambridge Infinitus Research Centre, University of Cambridge, Cambridge, UK. ⁵Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, UK. ⁶Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK. ⁷Institute of Human Genetics, University Medical Centre of Göttingen, Göttingen, Germany. ⁸Present address: Department of Molecular and Cellular Physiology and Howard Hughes Medical Institute, Stanford University Medical School, Stanford, CA, USA. *e-mail: fiona.watt@kcl.ac.uk

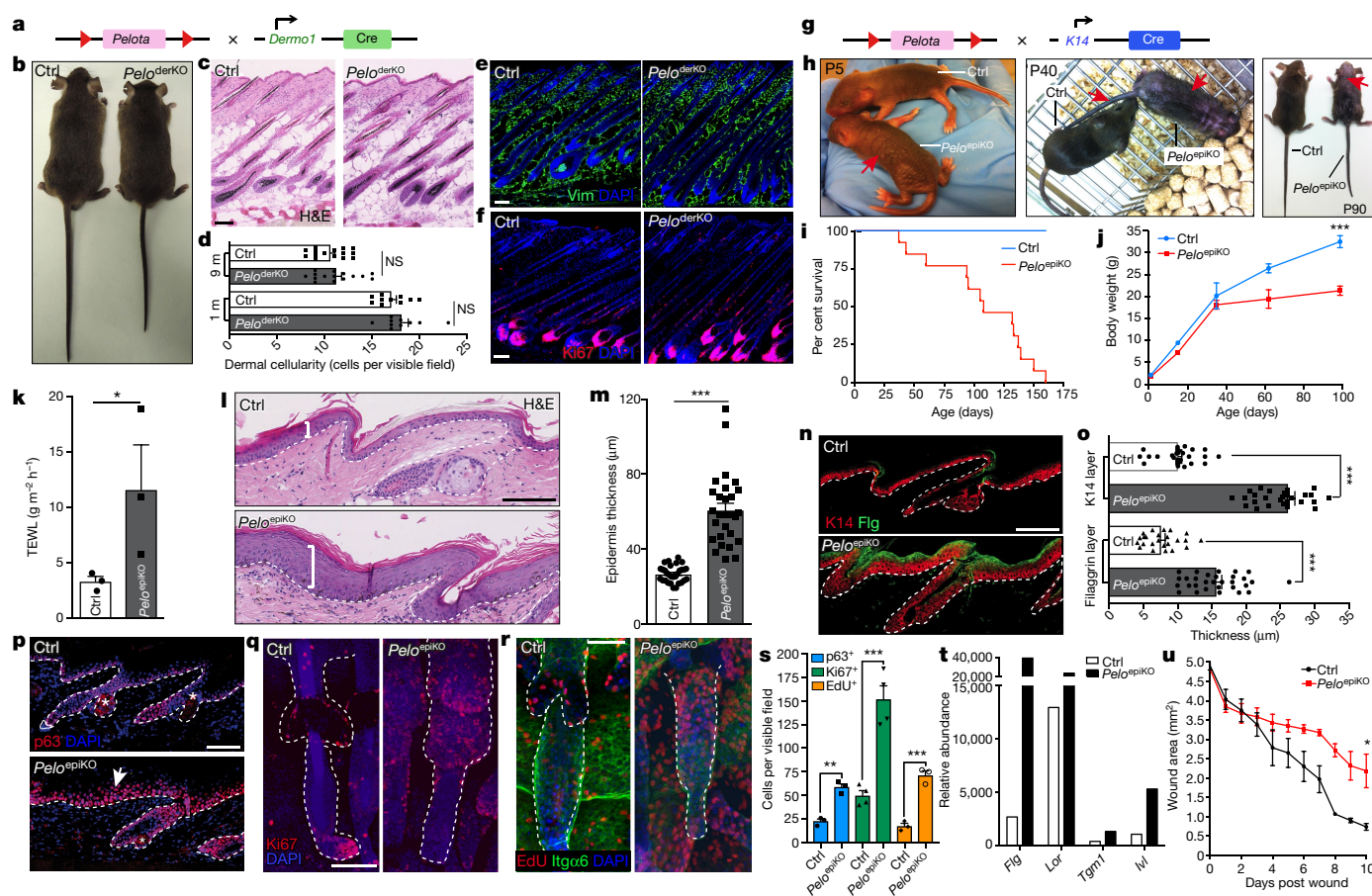


Fig. 1 | Differential effects of *Pelo* deletion. **a–f**, *PeloderKO* mice; **g–u**, *PeloeplKO* mice. **h**, Arrows show skin abnormalities. **c, d, l, m**, Haematoxylin and eosin (H&E) staining of back (**c, d**) and tail (**l, m**) skin. Dermal cellularity (**d**) and epidermal thickness (**m**) were measured. 1 m and 9 m, 1 and 9 months old, respectively. $n = 12$ sections analysed over three mice per group. **e, f, n, p–r**, Immunolabelling of sections (**e, f, n, p**) and wholemounts (**q, r**). Asterisks, non-specific; arrow, suprabasal labelling; dashed lines, epidermal–dermal boundary. **m, o**, *** $P < 0.001$, $n = 3$ mice.

i, Kaplan–Meier curves ($n = 29$ mice). **j**, Body weight. *** $P < 0.0003$; $n = 5$ per group. **k**, TEWL. $P < 0.05$; $n = 3$. **s**, Quantification of proliferation. ** $P = 0.0086$; *** $P = 0.0003$ for Ki67; *** $P = 0.0006$ for EdU; $n = 3$. **t**, Cumulative mean values of gene expression from ribosome profiling. **u**, Wound closure. * $P = 0.0500$; $n = 3$. Representative images in **c, e, f, l, n, p–r** from three independent experiments. Ctrl, littermate controls. Scale bars, 100 μm .

homologue of Rli1) rescue non-translating 3' UTR ribosomes¹³ and promote mRNA decay¹⁴. When we performed ribosomal profiling on keratinocytes from adult *PeloeplKO* mice by deep sequencing RPFs¹⁵, RPFs mapped primarily to the coding sequence (CDS) (Fig. 3a, Extended Data Fig. 7a, b), consistent with studies¹² showing that loss of PELO alone does not substantially increase 3' UTR ribosomes. CDS RPFs were primarily 28–34 nucleotides long, the expected fragment size protected by mammalian ribosomes¹⁶, and displayed the three-nucleotide periodicity that reflects codon-by-codon movement of elongating ribosomes (Fig. 3b, grey bars).

PeloeplKO profiles were enriched in 20–21-nucleotide RPFs (about 4–5% of total RPFs compared to less than 1% in control cells) (Fig. 3a–c). Like the dominant population of 28–34-nucleotide RPFs, these footprints were primarily found in the CDS and showed a strong reading frame signal, indicating that they too reflect the presence of elongating ribosomes, yet are shortened on their 3' ends after nuclease digestion (Fig. 3d, right). The density of short RPFs was evenly distributed and did not increase near the downstream 3' portion of transcripts (Fig. 3a), as would be anticipated if they resulted from ribosomes encountering a directional RNA decay process^{17,18}. Consistent with this, enrichment for 20–21-nucleotide footprints was not linked to reduced transcript abundance in *PeloeplKO* cells (Fig. 3e, Supplementary Table 1). Although *Pelo* has been implicated in the decay of unusual histone mRNAs that lack polyA tails¹⁹, the short footprints did not demonstrate patterns to indicate they result from ribosomes occupying transcripts that are being degraded. The 21mer RPFs seen in *PeloeplKO* cells could be the

equivalent of the 16mer species in yeast⁵ and reflect the increased size of the mammalian ribosome²⁰. However, we suggest they are equivalent to the 21-nucleotide fragments observed²¹ in anisomycin-treated yeast cells and reflect dependence on *Pelo*-associated quality control mechanisms in response to transfer RNA starvation in rapidly dividing cells.

Epidermal *Pelo* loss led to significant changes in global translational efficiency¹⁵ (Fig. 3f, g; $P < 0.01$). Translational efficiency values for keratins and ribosomal proteins were notably increased (Fig. 3e, f). There was substantial enrichment for genes involved in RNA metabolism, protein synthesis, extracellular matrix and chromatin regulation (Fig. 3h, Extended Data Fig. 7c–e, Supplementary Tables 2, 3). There was also differential expression of canonical translational pathways, including upregulation of the mTOR pathway (Fig. 3h, Extended Data Fig. 8a, b). As mTOR signalling leads to increased global translation²² (Extended Data Fig. 8c), we compared the previously published Gtpbp2/tRNA mutant⁸ and our *PeloeplKO* mouse cell gene expression datasets. We found substantial overlap in translational signalling pathways (Extended Data Fig. 8d), suggesting that ribosome stalling is sensed by mTOR.

The polysome-to-monomosome ratio was increased in *PeloeplKO* cells (Fig. 3i), suggesting an overall increase in translation or accumulation of inactive stalled ribosomes. *Krt86* transcripts were enriched in the heavy polysome fractions (Fig. 3j), consistent with the increases in translational efficiency values, suggesting increased overall translation. This was confirmed by quantifying global protein synthesis using O-propargyl-puromycin (OP-P) incorporation into newly

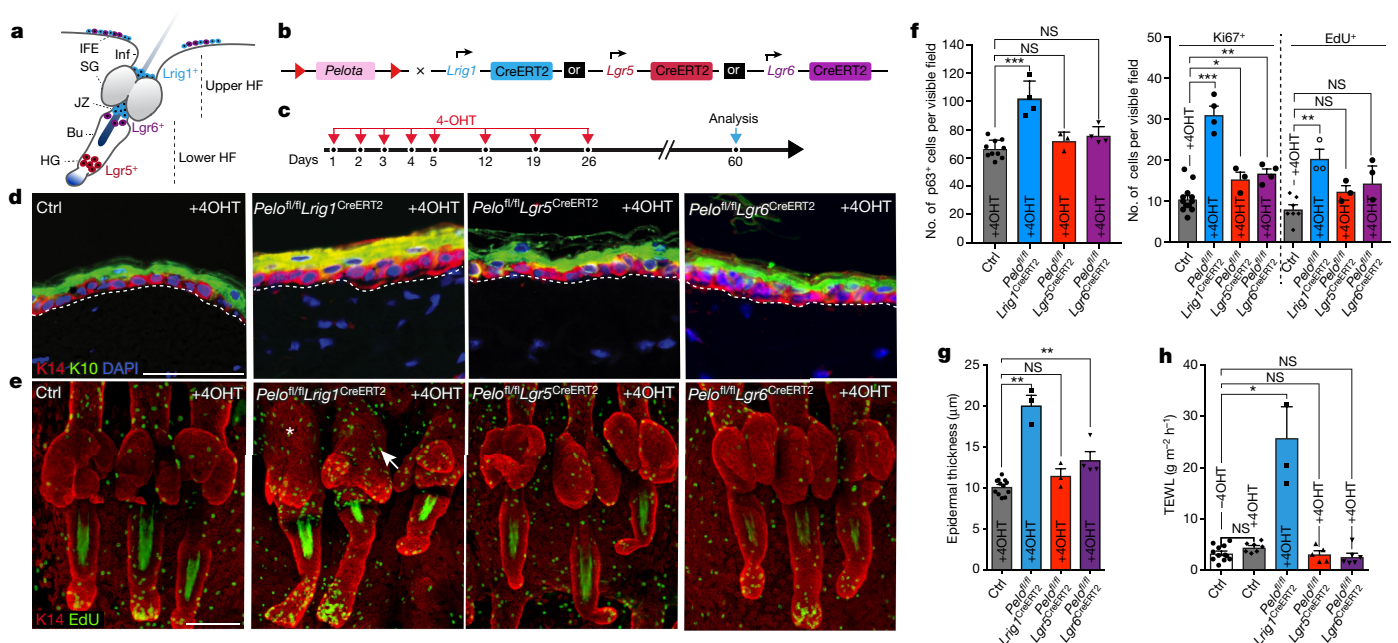


Fig. 2 | *Lrig1*⁺ stem cells account for *Pello* mutant epidermal phenotype. **a–c**, Schematics of *Lrig1*, *Lgr5* and *Lgr6* expression (**a**), breeding (**b**) and 4-OHT treatment (**c**). **d**, **e**, Immunostaining of dorsal skin IFE sections (**d**) and tail wholemounts (**e**) with antibodies against the markers shown. In **e**, asterisk, altered sebaceous gland; arrow, altered junctional zone. **f–h**, Quantification of proliferation (**f**), epidermal thickness (**g**) and TEWL (**h**). IFE, interfollicular epidermis; Inf, infundibulum; SG, sebaceous gland;

JZ junctional zone; Bu, bulge; HF, hair follicle; HG, hair germ. Scale bars, 50 μ m (**d**, **f**); 100 μ m (**e**). Dashed lines, epidermal–dermal boundary. *** P = 0.0010 (**f**, p63); *** P = 0.0005, * P = 0.0330, ** P = 0.0071 (**f**, Ki67); ** P = 0.0083 (**f**, EdU); ** P = 0.0044, 0.0011 (**g**); * P = 0.0167 (**h**). n = 16 sections and wholemounts analysed over four mice per group. NS, not significant.

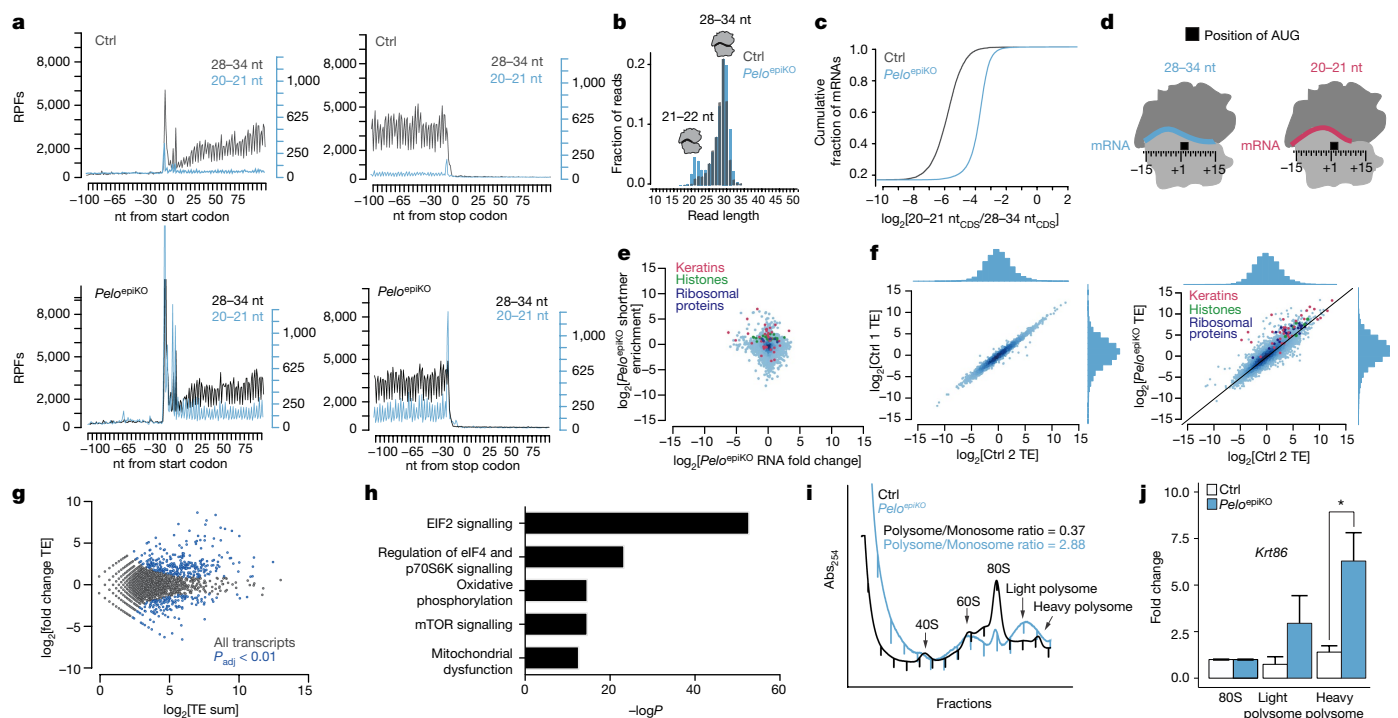


Fig. 3 | Accumulation of short ribosome footprints and global translational changes in *Pello* knockout epidermis. **a**, Metagenesis of full-length and short RPFs near the start (left) and stop (right) codons. **b**, RPF read length distributions. **c**, Empirical cumulative distribution plot of global enrichment of short 20–21-nucleotide (nt) reads relative to expected 28–34-nucleotide reads. **d**, Designations of -15 peaks indicate positions of 5' end of RPF; corresponding P site occupancy shown. **e**, Relative enrichment of short RPFs (y axis) and change in RNA transcript levels

(x axis). **f**, Replicate analysis of translational efficiency (TE). **g**, Average versus mean-difference (MA) plot showing observed and expected variance in translational efficiency measurements; adjusted P < 0.01, blue transcripts. **h**, Canonical pathways linked to translation regulation in *Pello*^{epiKO} cells. **i**, Epidermal polysome profiling. **j**, Quantitative PCR with reverse transcription (qRT-PCR) shows significant increase in heavy polysome-bound *Krt86* mRNA; P = 0.019.

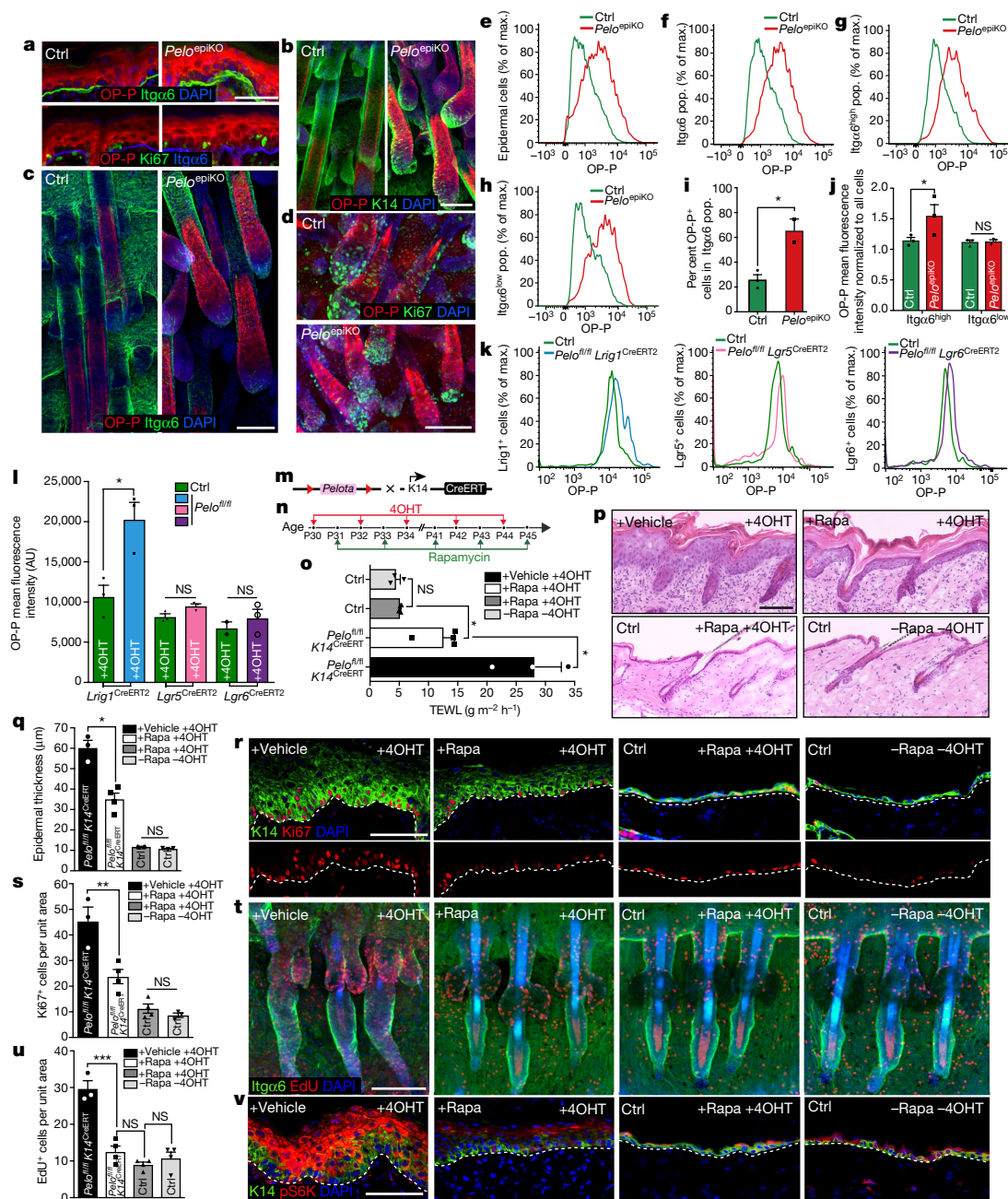


Fig. 4 | Inhibition of mTOR pathway attenuates *Pelo* phenotype progression. **a–d, r, t, v.** Immunolabelling for markers indicated. **s, u.** Quantification. $^{**}P = 0.0064$ (s); $^{***}P = 0.0006$ (u). **a–l.** Data from OP-P-injected newborn (a–j) and adult (k, l) mice. **e–k.** Representative flow histograms and quantification (i, j, l); $n = 3$ mice per group.

synthesized polypeptide chains^{3,4}. OP-P incorporation was increased in *Pelo*^{epiKO} IFE and hair follicles compared to controls. Labelling was higher in the IFE suprabasal layers than the basal layer, consistent with increased total protein synthesis during differentiation²³ (Fig. 4a–d). The increase in OP-P labelling in total *Pelo*-null keratinocytes and stem cells (Integrin $\alpha 6$ -high cells; Itga6^{high}) was confirmed by flow cytometry (Fig. 4e–j, Extended Data Fig. 9a). Confocal microscopy revealed a striking increase in the size of *Pelo*^{epiKO} basal cells (Extended Data Fig. 9b–d), consistent with increased protein synthesis and a higher proportion of G2/M and S phase cells (Extended Data Fig. 9e).

In control mice, Lgr1⁺ cells exhibited slightly higher protein synthesis than Lgr5⁺ and Lgr6⁺ cells (Fig. 4k, l). When *Pelo* was deleted, protein synthesis in Lgr1⁺ cells was increased further relative to Lgr5⁺ and Lgr6⁺ cells (Fig. 4k, l). RNA sequencing (RNA-seq) (Extended Data

$^{*}P = 0.0406$ (i), 0.0357 (j), 0.0198 (l). **m–v.** 4-OHT and rapamycin (Rapa) treatment. **o.** TEWL. $^{*}P = 0.0145$. **p, q.** Haematoxylin and eosin-stained dorsal skin. $^{*}P = 0.0286$. Scale bars, 50 μm (a); 100 μm (b–d, p, r, t, v); $n = 12$ sections and wholemounts analysed over four mice per group.

Fig. 10a) revealed that, regardless of whether *Pelo* was expressed, Lgr5⁺ cells clustered separately from Lgr1⁺ and Lgr6⁺ cells, while the gene expression profiles of individual populations did not cluster based on *Pelo* expression (Extended Data Fig. 10b–j, Supplementary Tables 4, 5). Therefore, the *Pelo* epidermal phenotype primarily reflects increased translation, rather than expression of specific genes.

To downregulate mTOR1²², we applied rapamycin to adult *Pelo*^{epiKO} skin (Extended Data Fig. 9f, g). There was a significant ($P < 0.02$) reduction in Ki67⁺ cells compared to controls (Extended Data Fig. 9h–j). Phosphorylated ribosomal protein S6K (pS6K), a key substrate of mTOR²², was increased in *Pelo*^{epiKO} skin, and reduced by rapamycin (Extended Data Fig. 9k). However, rapamycin did not prevent disruption of hair follicle and sebaceous gland architecture (Extended Data Fig. 9h).

Simultaneous rapamycin treatment and *Pelo* deletion largely prevented *Pelo*-mediated disruption of epidermal homeostasis (Fig. 4m, n). TEWL, epidermal thickening and proliferation were substantially reduced (Fig. 4o–u, Extended Data Fig. 9l); pS6K labelling was reduced (Fig. 4v) and phosphorylation of another mTOR substrate, 4EBP1, was decreased (Extended Data Fig. 9m). Therefore, the epidermal *Pelo* deletion phenotype is largely attributable to increased protein translation.

Our results indicate that translational control is critical for tissue homeostasis^{3,4,13} and establish a link between *Pelo* inactivation and translational activation via mTOR. mTOR is known to regulate cell growth and proliferation^{22,24} and is activated upon ribosome-stalling by Fragile X mental retardation protein^{25,26}. Impaired ribosomal biogenesis also activates mTOR1 signalling and stimulates translation initiation and elongation factors²⁷. mTOR signalling may be activated to enhance the efficiency of the translational machinery in order to compensate for impaired or reduced availability of ribosomes^{8,28}.

The increased size of *Pelo*-null epidermal cells as a result of increased protein synthesis^{23,29} may stimulate differentiation through decreased basement membrane engagement³⁰ and thus indirectly promotes proliferation. Factors that may account for the selective sensitivity of Lrig1⁺ cells to *Pelo* deletion include their proliferative state, abundance and location relative to Lgr5⁺ and Lgr6⁺ cells, together with their known ability to repopulate different epidermal compartments³¹.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0032-3>.

Received: 14 January 2017; Accepted: 28 February 2018;

Published online 11 April 2018.

- Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nat. Struct. Mol. Biol.* **23**, 7–15 (2016).
- Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* **19**, 594–601 (2012).
- Blanco, S. et al. Stem cell function and stress response are controlled by protein synthesis. *Nature* **534**, 335–340 (2016).
- Signer, R. A., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49–54 (2014).
- Guydosh, N. R. & Green, R. Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**, 950–962 (2014).
- Rezza, A. et al. Signaling networks among stem cell precursors, transit-amplifying progenitors, and their niche in developing hair follicles. *Cell Rep.* **14**, 3001–3018 (2016).
- Becker, T. et al. Structure of the no-go mRNA decay complex Dom34–Hbs1 bound to a stalled 80S ribosome. *Nat. Struct. Mol. Biol.* **18**, 715–720 (2011).
- Ishimura, R. et al. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* **345**, 455–459 (2014).
- Kretschmar, K., Weber, C., Driskell, R. R., Calonje, E. & Watt, F. M. Compartmentalized epidermal activation of β -catenin differentially affects lineage reprogramming and underlies tumor heterogeneity. *Cell Rep.* **14**, 269–281 (2016).
- Elkenani, M. et al. Pelota regulates epidermal differentiation by modulating BMP and PI3K/AKT signaling pathways. *J. Invest. Dermatol.* **136**, 1664–1671 (2016).
- Florin, L. et al. Delayed wound healing and epidermal hyperproliferation in mice lacking JunB in the skin. *J. Invest. Dermatol.* **126**, 902–911 (2006).
- Young, D. J., Guydosh, N. R., Zhang, F., Hinnebusch, A. G. & Green, R. Rli1/ABCE1 recycles terminating ribosomes and controls translation reinitiation in 3'UTRs *in vivo*. *Cell* **162**, 872–884 (2015).
- Mills, E. W., Wangen, J., Green, R. & Ingolia, N. T. Dynamic regulation of a ribosome rescue pathway in erythroid cells and platelets. *Cell Rep.* **17**, 1–10 (2016).
- Mills, E. W., Green, R. & Ingolia, N. T. Slowed decay of mRNAs enhances platelet specific translation. *Blood* **129**, e38–e48 (2017).
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).
- Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
- Guydosh, N. R., Kimmig, P., Walter, P. & Green, R. Regulated Ire1-dependent mRNA decay requires no-go mRNA degradation to maintain endoplasmic reticulum homeostasis in *S. pombe*. *eLife* **6**, e29216 (2017).
- Guydosh, N. R. & Green, R. Translation of poly(A) tails leads to precise mRNA cleavage. *RNA* **23**, 749–761 (2017).
- Slevin, M. K. et al. Deep sequencing shows multiple oligouridylations are required for 3' to 5' degradation of histone mRNAs on polyribosomes. *Mol. Cell* **53**, 1020–1030 (2014).
- Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209 (2010).
- Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**, e01257 (2014).
- Thoreen, C. C. et al. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
- Kristensen, A. R., Gsponer, J. & Foster, L. J. Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* **9**, 689 (2013).
- Laplane, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
- Darnell, J. C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Sawicka, K., Pyronneau, A., Chao, M., Bennett, M. V. & Zukin, R. S. Elevated ERK/p90 ribosomal S6 kinase activity underlies audiogenic seizure susceptibility in fragile X mice. *Proc. Natl Acad. Sci. USA* **113**, E6290–E6297 (2016).
- Liu, R. et al. Impairing the production of ribosomal RNA activates mammalian target of rapamycin complex 1 signalling and downstream translation factors. *Nucleic Acids Res.* **42**, 5083–5096 (2014).
- Mills, E. W. & Green, R. Ribosomopathies: There's strength in numbers. *Science* **358**, eaan2755 (2017).
- Gingold, H. et al. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281–1292 (2014).
- Watt, F. M., Jordan, P. W. & O'Neill, C. H. Cell shape controls terminal differentiation of human epidermal keratinocytes. *Proc. Natl Acad. Sci. USA* **85**, 5576–5580 (1988).
- Page, M. E., Lombard, P., Ng, F., Götting, B. & Jensen, K. B. The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell* **13**, 471–482 (2013).

Acknowledgements K. L.-A. thanks W. Engel for inspiration. F.M.W. acknowledges funding from the Wellcome Trust, Cancer Research UK and the UK Medical Research Council. We are also grateful for funding from the Department of Health via the National Institute for Health Research comprehensive Biomedical Research Centre award to Guy's & St Thomas' National Health Service Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. R.G. acknowledges funding from Howard Hughes Medical Institute. K.H.S. is the recipient of a Finnish Cultural Foundation Fellowship and T.L. is supported by a Sir Henry Dale Fellowship (206211/Z/17/Z). We thank C. Lelliott and V. Vancollie for *Hbs1* knockout samples, and A. Ahmed, G. Donati, R. Driskell, S. Habib, M. Lynch, B. Oules, D. V. K. Pantakani, I. Sancho-Martinez, G. Walko and all members of the Watt laboratory for discussions. Technical help from M. Battilocchi, C. Li and staff at the Nikon Imaging Centre and the Hodgkin Biological Services Facility, King's College London, is gratefully acknowledged. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, Oxford, UK (funded by Wellcome Trust: 203141/Z/16/Z) for generation of the sequencing data.

Reviewer information Nature thanks X. Dai, V. Horsley and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.L.-A. and F.M.W. conceptualized and led the study. K.L.-A., I.S. and B.M.L. performed and analysed mouse experiments. K.L.-A., K.H.S. and A.M. performed and analysed cell culture experiments. A.O.P. analysed data from ribosome profile and RNA-seq experiments. I.M.A. generated the *Pelo* conditional knockout mouse. E.W.M., C.C.-C.W. and R.G. generated and analysed ribosome-profiling data. H.Y., T.L. and A.I.L. generated and analysed polysome data. K.L.-A. and F.M.W. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0032-3>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0032-3>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to F.M.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mouse strains. All mouse experiments were performed under a UK Government Home Office project license and subject to local institutional ethical approval. The generation of conditional *Pelo*^{fl/fl} (*Pelo*^{tm1lmad}) mice has been described elsewhere³². To derive constitutive *Pelo* epidermal knockout mice (*Pelo*^{epiKO}), *Pelo*^{fl/fl} mice were crossed with *Krt14*^{Cre} mice (Jax strain, stock number 004782). To achieve temporally controlled *Pelo* knockout and genetic labelling of cells lacking *Pelo*, *Pelo*^{fl/fl} mice were crossed with *Krt14*^{CreERT} (Jax strain, stock number 005107), *Lrig1*^{EGFP-IRES-CreERT2} mice³¹, *Lgr5*^{EGFP-IRES-CreERT2} mice³³, *Lgr6*^{EGFP-IRES-CreERT2} mice³⁴ and *Rosa26*^{LoxP-Stop-LoxP-tdTomato} mice³⁵. To activate Cre recombinase, 4-OHT (Sigma-Aldrich) was dissolved in acetone and applied topically (3 mg/100 µl) every day for five days and once a week for three weeks. For proliferation assays, 5-ethynyl-2'-deoxyuridine (EdU) (Invitrogen, 20 mg per kg body mass; in PBS) was injected intraperitoneally and the tissue was removed 1 h later. To derive constitutive *Pelo* dermal knockout mice (*Pelo*^{derKO}), *Pelo*^{fl/fl} mice were crossed with *Dermo1*^{Cre} (B6.129 × 1-Twist2^{tm1.1(cre)Dor/J}) mice^{36,37}. Mouse lines used in this study and the locations of marker expression in the skin are illustrated in Extended Data Fig. 10k. *Hbs1l*^{-/-} (*Hbs1l*^{tm1a(KOMP)Wtsi}) mice were produced at the Wellcome Trust Sanger Institute Mouse Genetics Project as part of the International Mouse Phenotype Consortium (IMPC)³⁸.

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Library generation for ribosome profiling. Samples of *Pelo*^{epiKO} epidermis for ribosome profiling and RNA-seq were prepared by scraping off the epidermal layer in liquid nitrogen. Frozen samples were ground using a Mixer Mill (Retsch) and thawed in the presence of polysome lysis buffer. Lysates were clarified by centrifugation at 20,000g for 10 min at 4 °C and the supernatant was collected. Total lysate RNA was quantified using the Quant-it RNA kit (Thermo) and 5 µg was used for preparation of ribosome profiling libraries as described previously¹⁵. Total RNA was size-selected by excising gel regions between phosphorylated 16-nt and 34-nt RNA oligo standards. Ribosomal RNAs were depleted using Ribo-Zero Gold (Illumina) after footprint size-selection. One hundred nanograms ribosomal RNA was used for preparation of RNA-seq libraries from the same samples as profiling libraries. Analysis using a BioAnalyzer total RNA pico chip was used to confirm RNA integrity (RIN > 9) for RNA sequencing samples. The datasets are deposited in GEO under accession number GSE94385.

Sequencing and data analysis. Ribosome profiling and RNA-seq libraries were sequenced using a HiSeq2500 (Illumina). About 110 million total raw reads were generated from 4 ribosome profiling samples with between 11 and 30 million reads mapping to the genome per sample. For ribosome profiling analysis, only singly mapped reads (NH:i:1) with no mismatches (NM:i:0) were used. Translational efficiency (TE) was calculated as the number of CDS RPFs per RPKM. Relative 3' UTR ribosome occupancy was calculated as 3' UTR footprint density/CDS footprint density. For differential gene expression analysis, we uploaded the list of differentially expressed genes into Ingenuity IPA and ran a core analysis. This identified the top molecules, pathways and master regulators that differed between control and *Pelo*^{epiKO} samples.

Polysome analysis. Epidermal layers from wild-type and *Pelo*^{epiKO} mice were lysed as described in 'Library generation for ribosome profiling'. Clarified lysates were loaded on 10–50% sucrose gradients prepared in polysome gradient buffer (20 mM Tris-HCl pH 8, 150 mM KCl, 5 mM MgCl₂, 0.5 mM DTT, 0.1 mg/ml cycloheximide), and gradients were spun in an SW41-Ti rotor at 40,000 r.p.m. for 3 h at 4 °C. Gradients were fractionated using a Brandel Density Gradient Fractionation System. Prior to RNA extraction, CLuc mRNA (NEB) was added to each fraction. RNA was extracted using hot acidic phenol and cDNA was synthesized using iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. qPCR was carried out using iTaq Universal SYBR Green Supermix (Bio-Rad). Relative mRNA abundances in indicated fractions were normalized to CLuc mRNA to account for differences in RNA extraction efficiency among fractions, and then calculated as fold changes normalized to 80 S fractions. qPCR primers: *CLuc* forward 5'-GCTTCAACATCACCGTCATTG-3', *CLuc* reverse 5'-CACAGAGGCCAGAGATCATTC-3', *Krt86* forward 5'-AACA GAATGATCCAGAGGCTG-3', *Krt86* reverse 5'-GCTCAGATGGGTCACGG-3'.

RNA-seq library preparation and analysis. A primary epidermal cell suspension was prepared as previously described³⁹. In brief, cells were harvested from 3-month-old 4-OHT-treated *Pelo*^{fl/+} *Lrig1*^{EGFP-CreERT2}, *Pelo*^{fl/+} *Lgr5*^{EGFP-CreERT2} and *Pelo*^{fl/+} *Lgr6*^{EGFP-CreERT2} control mice, and *Pelo*^{fl/fl} *Lrig1*^{EGFP-CreERT2}, *Pelo*^{fl/fl} *Lgr5*^{EGFP-CreERT2} and *Pelo*^{fl/fl} *Lgr6*^{EGFP-CreERT2} *Pelo* mutant mice. The total epidermal population was sorted by fluorescence-activated cell sorting (FACS) for GFP⁺ cells on a BD FACSARIAII cell Sorter and 1,000 GFP-high cells collected from each population for RNA-seq. Library construction and the strategy for RNA-seq involved the Smart-seq2 method as reported previously⁴⁰. Fastq files of paired-end reads were uploaded to the Galaxy platform⁴¹ and aligned using STAR aligner⁴²

to the *Mus musculus* reference genome (GRCm38/Mm10). BAM files were processed in R using the 'rseqGene' workflow⁴³. The data were analysed using the edgeR package. Processed data were mined using IPA Ingenuity Pathway Analysis (Qiagen). The datasets are deposited in GEO under accession number GSE106246.

Flow cytometry for measurement of cell size, cell cycle and protein synthesis in vivo. To analyse cell size by flow cytometry, epidermal cells were isolated as previously described³⁹. In brief, epidermis was enzymatically separated from dermis with thermolysin (Sigma, 0.25 mg/ml in PBS) overnight at 4 °C. Epidermal sheets were processed into single cell suspensions by incubation in DMEM (Gibco) containing DNase (Sigma, 250 µg/ml) for 20 min at 37 °C with shaking. Single cells were labelled according to standard procedures with anti-Integrin α6-Alexa Fluor 647 or FITC (AbSource, 1:20) antibodies. To assess the percentage of proliferating epidermal cells, mice were injected with 500 µg 5-ethynyl-2'-deoxyuridine (EdU; 2.5 mg/ml in PBS) intraperitoneally and back skin was harvested 2 h later. Cells were isolated as described above and single cell suspensions were stained with the Click-iT EdU Alexa Fluor 488 Flow Cytometry Kit (Invitrogen) according to the manufacturer's recommendations. Cell cycle analysis was performed on a BD LSR Fortessa cell analyser. Proliferating cells that had incorporated EdU were detected in the FITC/Alexa Fluor 488 channel.

To measure protein synthesis in vivo, we injected mice intraperitoneally with OP-P (Medchem Source or Thermo Fisher (C10459); 50 mg kg⁻¹ body mass; pH 6.4–6.6 in PBS). One hour later, mice were euthanized and back and tail skin samples were collected. Epidermal dissociation was performed as described above. The staining for detection of protein synthesis was performed according to the manufacturer's instructions (Click-iT Plus OPP Protein Synthesis Assay Kit; ThermoFisher Scientific). Samples from PBS-injected mice were also stained for detection of protein synthesis and the fluorescence signal was used to determine background labelling. Rates of protein synthesis were calculated as described previously³. In brief, OP-P signals were normalized to whole epidermis after subtracting the autofluorescence background. 'Mean OP-P fluorescence' reflected fluorescence values for each cell population normalized to whole epidermis. Labelled cells were analysed on a BD LSRFortessa cell analyser. All data were analysed using FlowJo software.

Histology, epidermal wholemounts and imaging. For paraffin sections, skin samples were fixed with 10% neutral buffered formalin overnight before paraffin embedding. The tissues were sectioned and stained with H&E and Herovici's stain by conventional methods. For frozen sections, skin samples were embedded in OCT (optimal cutting temperature compound; VWR), sectioned and fixed in 4% PFA for 10 min before staining. Slides were mounted using ProLong Gold anti-fade reagent containing DAPI (Life Technologies) as a nuclear counterstain. H&E and Herovici images were acquired using a Hamamatsu slide scanner and analysed using NanoZoomer software (Hamamatsu).

The epidermal wholemount labelling procedure was performed as described previously^{44,45}. In brief, mouse tails were slit on the ventral side lengthways. Pieces (0.5 × 0.5 cm²) of skin were incubated in 5 mM EDTA in PBS at 37 °C for 4 h. The epidermis was gently peeled from the dermis as an intact sheet in a proximal to distal direction, corresponding to the orientation of the hairs, and then the epidermis was fixed in 4% paraformaldehyde (PFA; Sigma) for 1 h at room temperature. Fixed epidermal sheets were washed in PBS and stored in PBS containing 0.2% sodium azide at 4 °C.

Confocal image acquisition of stained wholemounts and skin sections was performed using a Nikon A1 confocal microscope. Images were analysed using NIS Elements (Nikon Instruments Inc.). Photoshop CS5 (Adobe image suite) was used to optimize the images globally for brightness, contrast and colour balance.

Rapamycin treatment. Rapamycin (LC Laboratories, R5000) was dissolved in acetone. Rapamycin treatment groups of mice received topical applications of 500 µl 0.2% rapamycin on dorsal and tail skin. Vehicle treatment groups received an equal volume of acetone without rapamycin. Dorsal skin was shaved before the day of treatment.

Wound and TEWL assays. Full-thickness wounds were made on the lower dorsal skin (5 mm) or tail (2 mm) using a punch biopsy (Stiefel) under analgesia and general anaesthesia. The hair on the back was shaved before wounding. Wound closure was measured using a Vernier scale. Epidermal barrier function was assessed by testing basal TEWL on the dorsal skin of mice using a TEWAmetre (Courage and Khazaka, TM210). Measurements were collected for 15–20 s when TEWL readings had stabilized, approximately 30 s after the probe collar was placed on the dorsal skin.

Antibodies. Primary antibodies for wholemount and tissue sections were: chicken anti-Krt14 (Covance, SIG2376, 1:500) or directly conjugated (AlexaFluor 555) Krt14 (LL002, in house, 1:200); directly conjugated (AlexaFluor 488) Krt15 (LHK-15, in-house, 1:50); rabbit anti-p63 (SCBT, sc367333, 1:100); rabbit anti-filaggrin (Covance, PRB-417P, 1:100); mouse anti-FASN (SCBT, sc48357, 1:100); rabbit anti-Ki67 (Novocastra, NCL-Ki67p, 1:500); rabbit anti-Ki67 (Abcam, ab16667, 1:500); rabbit anti-phospho-S6 ribosomal protein (Ser235/236) (pS6K, Cell

Signaling, 2211, 1:200); rabbit anti-P-cadherin (Cell Signaling, 2130, 1:200); rabbit anti-Vimentin (Cell Signaling, 5741s, 1:500); rabbit anti-Krt10 (Covance, PRB-159P, 1:500); FITC-conjugated rat anti-CD49f (Integrin $\alpha 6$, Biolegend, 313606, 1:100); goat anti-Lrig1 (R&D Systems, FAB3688G, 1:200); rabbit anti-Scd1 (Cell Signaling, 2794s, 1:500); mouse anti-involucrin (SY5, in-house, 1:500); mouse anti-pan-keratin (Abcam, ab8068, 1:200); rat anti-CD34 (RAM34, Thermo Fisher, 14-0341-82, 1:200); rabbit anti-phospho-4EBP1 (Thr37/46) (Cell Signaling, 236B4, 1:500). AlexaFluor (Life Technologies) dye-conjugated secondary antibodies were used at 1:250 dilutions.

In vitro knockdown, clonogenicity and skin reconstitution assay. Primary human keratinocytes (strain km) were isolated from neonatal foreskin and cultured on mitotically inactivated 3T3-J2 feeder cells in complete FAD medium, containing 1 part Ham's F12 medium and three parts Dulbecco's modified Eagle's medium (DMEM), 1.8×10^{-4} M adenine, 10% (v/v) FBS, $0.5 \mu\text{g ml}^{-1}$ hydrocortisone, $5 \mu\text{g ml}^{-1}$ insulin, 10^{-10} M cholera toxin and 10 ng ml^{-1} EGF, as described previously^{46,47}. Keratinocytes routinely tested negative for mycoplasma but were not subjected to STR profiling because they were not an established cell line. siRNA-mediated gene silencing was performed as described previously⁴⁸. In brief, keratinocytes were transferred to feeder-free conditions in keratinocyte serum-free medium (KSFM) containing $30 \mu\text{g ml}^{-1}$ BPE (bovine pituitary extract) and 0.2 ng ml^{-1} EGF (Gibco) for 2–3 days. Cells were trypsinized at ~70% confluence and resuspended in cell line buffer SF (Lonza). For each $20\text{-}\mu\text{l}$ transfection (program FF-113), 2×10^5 cells were mixed with $1\text{--}2 \mu\text{M}$ siRNA duplexes (Silencer select siRNA for *PELO* ID131910, ID131911, ID131912, as well as negative control, Ambion). Transfected cells were incubated at room temperature for 5–10 min and subsequently resuspended in pre-warmed KSFM. siRNA nucleofections were performed with the Amaxa 16-well shuttle system (Lonza). Alternatively, keratinocytes were transfected by using INTERFERin (Polyplus transfections): 36 pmol siRNA , $4 \mu\text{l}$ INTERFERin reagent, and $200 \mu\text{l}$ KSFM were mixed in a collagen-coated ($20 \mu\text{g ml}^{-1}$ in PBS, 1 h, 37°C) 12-well plate and incubated for 20 min at room temperature. After the incubation, 75,000 keratinocytes were seeded per well (final concentration of siRNA 30 nM). Medium was changed after 4 h and cells were harvested after 48 h.

For clonogenicity assays, nucleofected keratinocytes were seeded at low density (100–250 cells per well) on a prepared feeder layer in 6-well plates containing complete FAD medium. Keratinocytes were maintained in culture for 12 days and then feeders were removed by Versene treatment combined with tapping the culture flask. Once all the feeder cells had been washed away, the remaining keratinocyte colonies were fixed with 4% PFA at room temperature for 10 min. Colonies were then stained with 1% Rhodanile Blue (1:1 mixture of Rhodamine B and Nile Blue A (Acros Organics) solution for 15 min and washed with distilled water before examination. Stained dishes containing keratinocyte colonies were imaged using a Molecular Imager Gel Doc XR + imaging system (Bio-Rad). Colonies were measured using ImageJ and clonogenicity was calculated as the percentage of plated cells that formed colonies.

For the skin reconstitution assay, pre-confluent keratinocyte cultures (km passage 3) were disaggregated and transfected either with *PELO* siRNAs or scrambled control siRNAs. Twenty-four hours after transfection, keratinocytes were collected and seeded on irradiated de-epidermized human dermis in 6-well Transwell plates with feeders and cultured at the air–liquid interface for three weeks⁴⁹. Organotypic cultures were fixed in 10% neutral buffered formalin (overnight), paraffin embedded and sectioned for H&E and immunofluorescence analysis.

Picrosirius birefringence, dermal thickness and dermal cell density. Twelve-micrometre paraffin sections were stained with picrosirius red using a standard method⁵⁰. In brief, the sections were de-paraffinized, washed twice with water and stained for 1 h in picrosirius red solution (0.1% Sirius red F3B in a saturated aqueous solution of picric acid). After staining, sections were washed twice with acidified water (0.5% acetic acid), dehydrated, cleared with xylene, and mounted

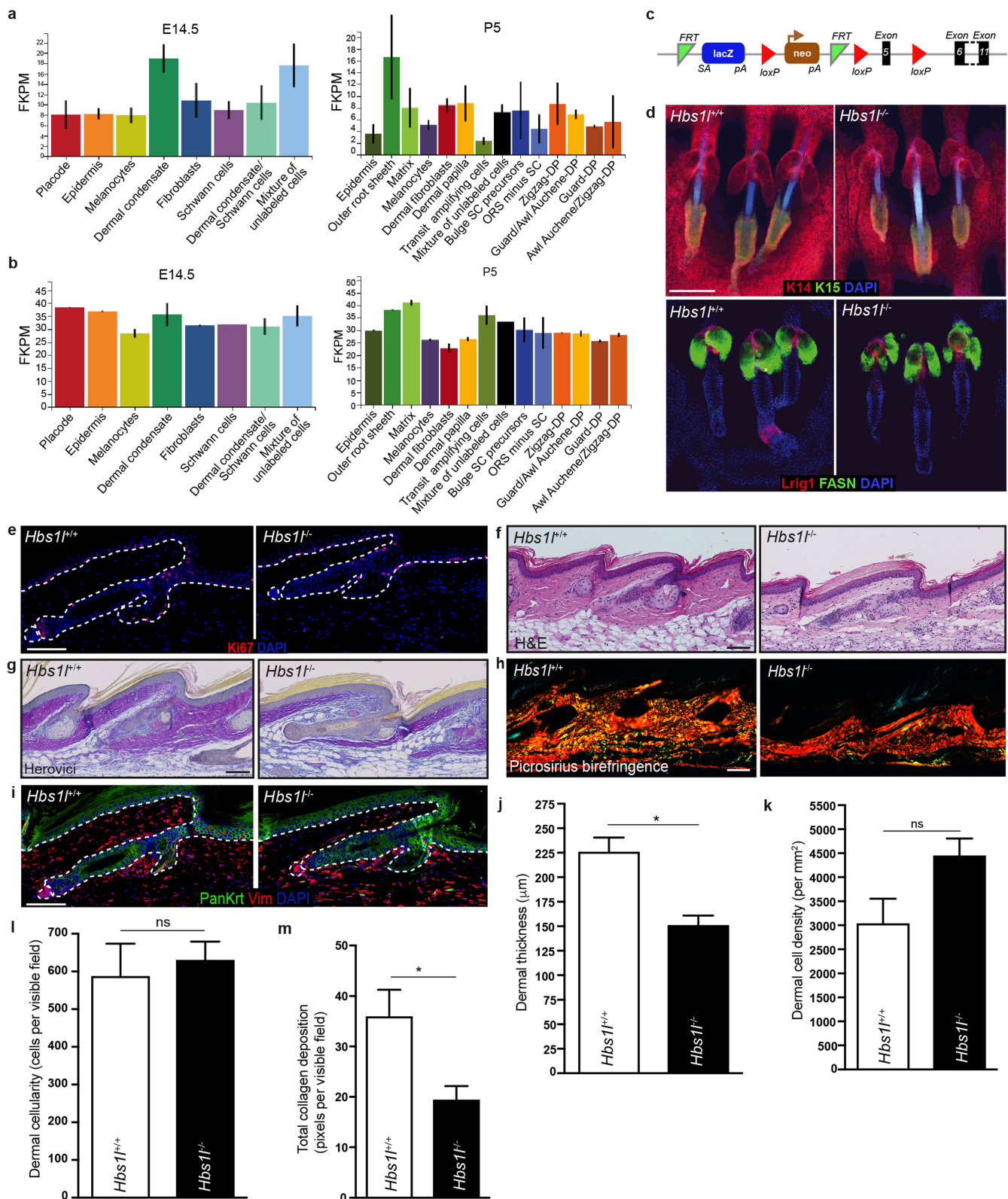
with DPX mounting medium. The images were acquired using a Zeiss Axiophot microscope and AxioCam HRC camera under plane-polarized light. The quantification of total collagen fibres was performed using Fiji (ImageJ) software. The collagen pixels were selected with the Colour Threshold tool (hue 0–100, saturation 0–255 and brightness 230–255). Thickness of dermis was quantified by NanoZoomer Digital Pathology software (Hamamatsu). The number of cells was determined with ImageJ by counting nuclei in DAPI stained tissue sections.

Statistics. Statistical significance in all experiments was calculated by Student's *t*-test. Data are represented as mean \pm s.e.m. (error bars). GraphPad Prism was used for calculation and illustration of graphs.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

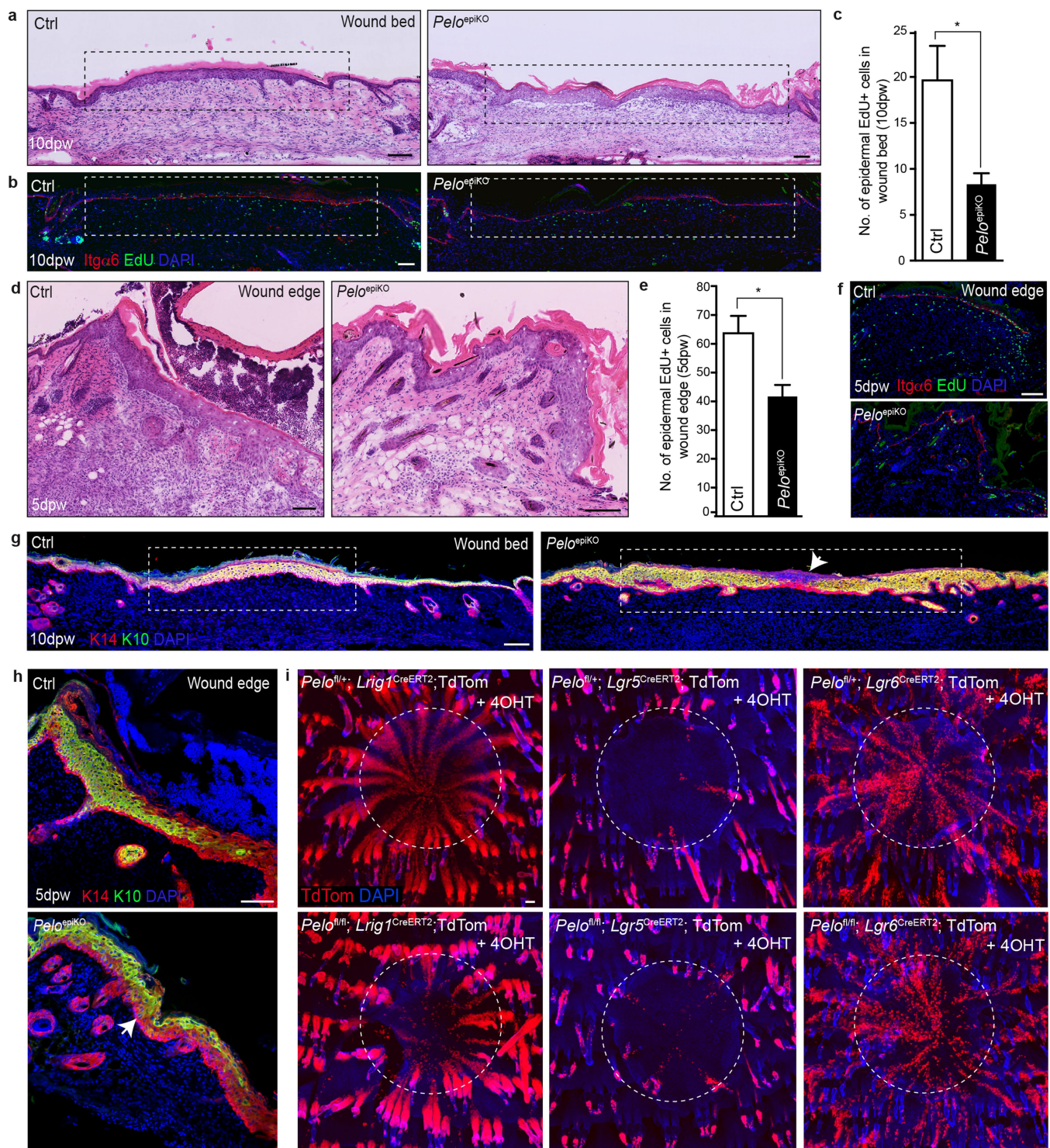
Data availability. All experimental data generated and/or analysed during this study are included in this published article (and its Supplementary Information files). In addition, ribosome profiling data (accession number GSE94385) and RNA-seq data (accession number GSE106246) are available in GEO.

32. Nyamsuren, G. et al. Pelota regulates the development of extraembryonic endoderm through activation of bone morphogenetic protein (BMP) signaling. *Stem Cell Res.* **13**, 61–74 (2014).
33. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
34. Snippert, H. J. et al. Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. *Science* **327**, 1385–1389 (2010).
35. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
36. Šošić, D., Richardson, J. A., Yu, K., Ornitz, D. M. & Olson, E. N. Twist regulates cytokine gene expression through a negative feedback loop that represses NF- κ B activity. *Cell* **112**, 169–180 (2003).
37. Lichtenberger, B. M., Mastrogiannaki, M. & Watt, F. M. Epidermal β -catenin activation remodels the dermis via paracrine signalling to distinct fibroblast lineages. *Nat. Commun.* **7**, 10537 (2016).
38. Skarnes, W. C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
39. Jensen, K. B., Driskell, R. R. & Watt, F. M. Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis. *Nat. Protoc.* **5**, 898–911 (2010).
40. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
41. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**(W1), W3–W10 (2016).
42. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
43. Love, M. I., Anders, S., Kim, V. & Huber, W. RNA-seq workflow: gene-level exploratory analysis and differential expression. *F1000 Res.* **4**, 1070 (2015).
44. Braun, K. M. et al. Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in whole mounts of mouse epidermis. *Development* **130**, 5241–5255 (2003).
45. Liakath-Ali, K. et al. Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen. *Nat. Commun.* **5**, 3540 (2014).
46. Lowell, S., Jones, P., Le Roux, I., Dunne, J. & Watt, F. M. Stimulation of human epidermal differentiation by delta-notch signalling at the boundaries of stem-cell clusters. *Curr. Biol.* **10**, 491–500 (2000).
47. Gandarillas, A. & Watt, F. M. Changes in expression of members of the fos and jun families and myc network during terminal differentiation of human keratinocytes. *Oncogene* **11**, 1403–1407 (1995).
48. Mulder, K. W. et al. Diverse epigenetic strategies interact to control epidermal differentiation. *Nat. Cell Biol.* **14**, 753–763 (2012).
49. Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463**, 563–567 (2010).
50. Lattouf, R. et al. Picrosirius red staining: a useful tool to appraise collagen networks in normal and pathological tissues. *J. Histochem. Cytochem.* **62**, 751–758 (2014).



Extended Data Fig. 1 | *Pelo* is expressed in all skin cell subpopulations and knockout of *Hbs11* leads to mild dermal phenotype. **a, b**, *Pelo* and *Hbs11* are ubiquitously expressed in all cell populations of embryonic and neonatal skin. mRNA expression data obtained from hair and skin gene expression library (Hair-GEL; <http://www.hair-gel.net>). **c**, Schematic of *Hbs11* knockout first allele. **d**, Immunolabelling of tail epidermal wholemounts with antibodies against Krt14, Krt15, Lrig1 and FASN. **e**, Tail skin sections immunolabelled for Ki67, showing no significant change in the distribution of Ki67⁺ cells in *Hbs11*^{-/-} epidermis. **f**, H&E

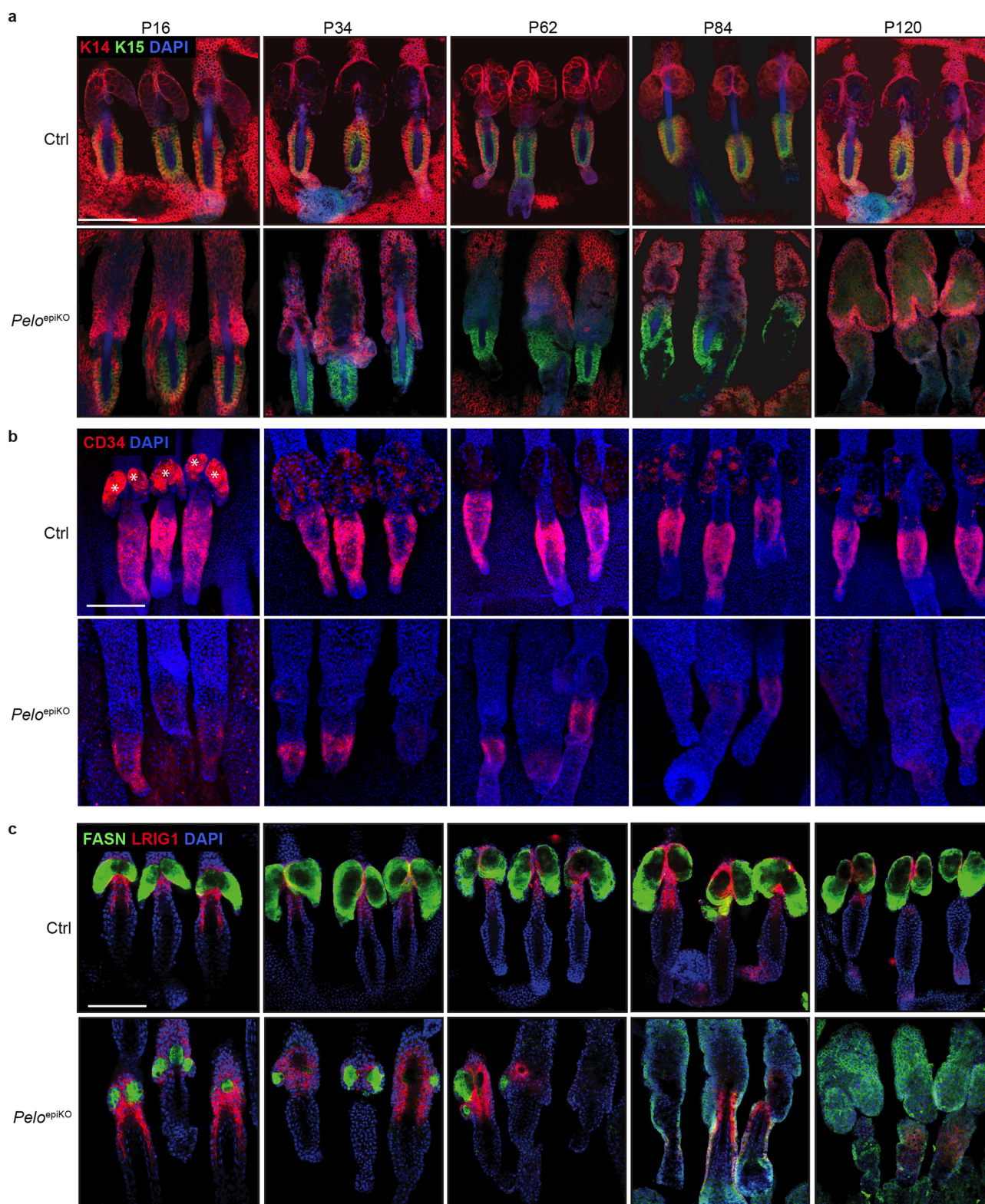
staining of adult control and *Hbs11*^{-/-} tail skin. **g**, Herovici's polychrome staining to visualize immature (blue) and mature (pink) dermal collagen. **h**, Picrosirius staining of tail skin showing the birefringence of collagen fibres against a black background. **i**, Immunostaining of tail skin sections with pan-keratin (PanKrt) and vimentin (Vim) antibodies. **j–m**, Quantification of dermal thickness (**j**), dermal cell density (**k**), dermal cellularity (**l**) and total collagen deposition (**m**). Dashed lines mark epidermal–dermal boundary. Scale bars, 100 μm. **P* = 0.0286 (**j, m**). *n* = 12 sections analysed over four mice per group.



Extended Data Fig. 2 | Delayed wound closure in *Pelo* null epidermis.

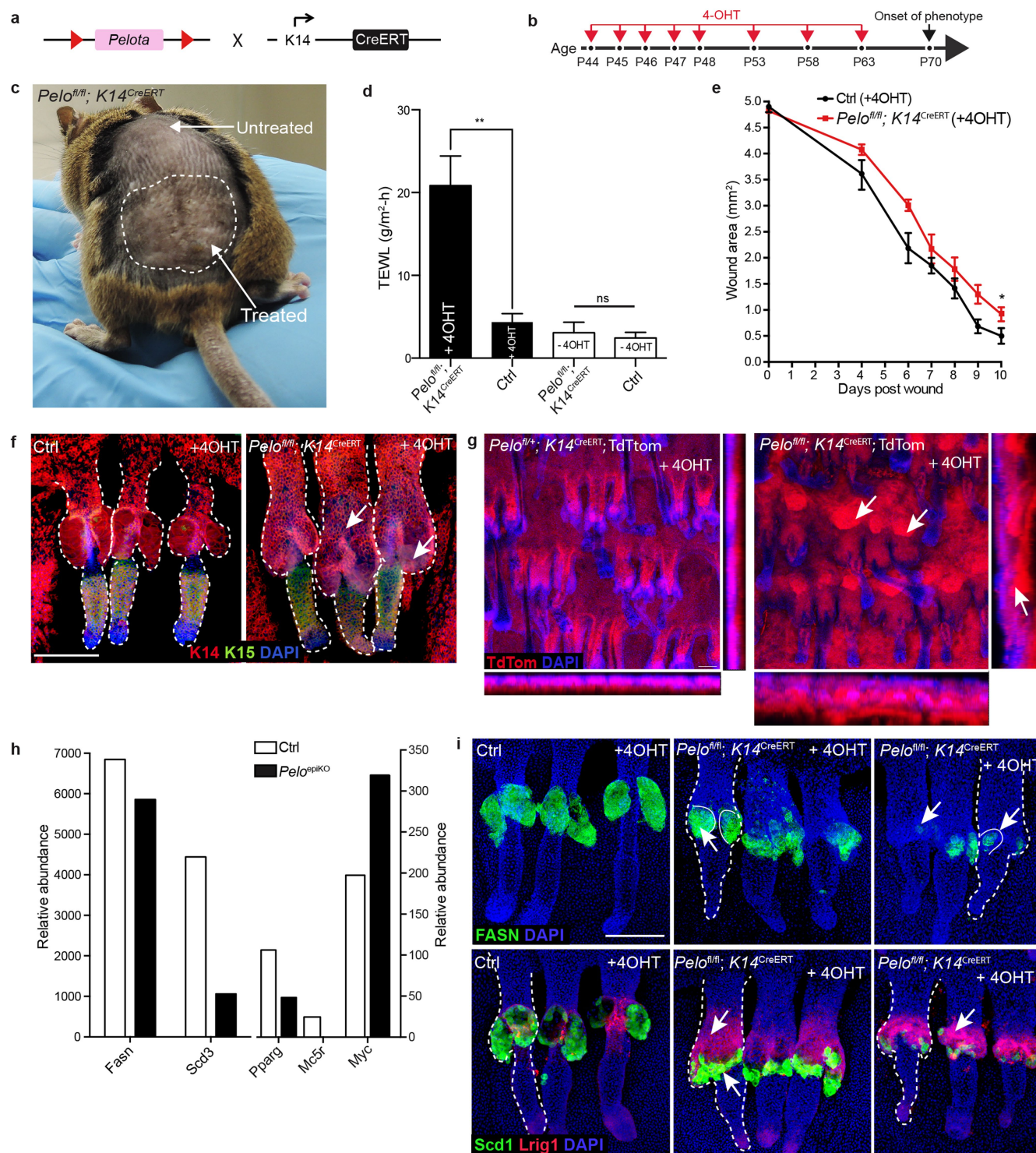
a, Histology of skin 10 days post wound (dpw) shows delayed wound closure in *Pelo*^{epiKO} mice. **b**, **c**, EdU staining of 10 dpw skin shows reduced proliferation in the wound bed. *Itga6* staining demarcates dermal–epidermal boundary. Box indicates the wound bed. **d**, Histology of 5 dpw wound shows altered epidermal architecture. **e**, **f**, EdU labelling of 5 dpw skin shows reduced proliferation at wound edge. **g**, **h**, Immunostaining

of Krt14 and Krt10 in 10 and 5 dpw skin shows abnormal differentiation in *Pelo*^{epiKO} mice (arrows). **i**, TdTomato genetic labelling shows the contribution of *Lrig1*, *Lgr5* and *Lgr6* progeny in tail wound healing. Note altered migration of *Lrig1* cells in *Pelo*^{fl/fl}*Lrig1*^{CreERT2} tdTomato mice when compared to *Lgr5* and *Lgr6* on *Pelo* deletion. **P* = 0.0123 (**c**), **P* = 0.0330 (**e**); *n* = 9 sections analysed over 3 mice per group. Scale bars, 100 μ m.



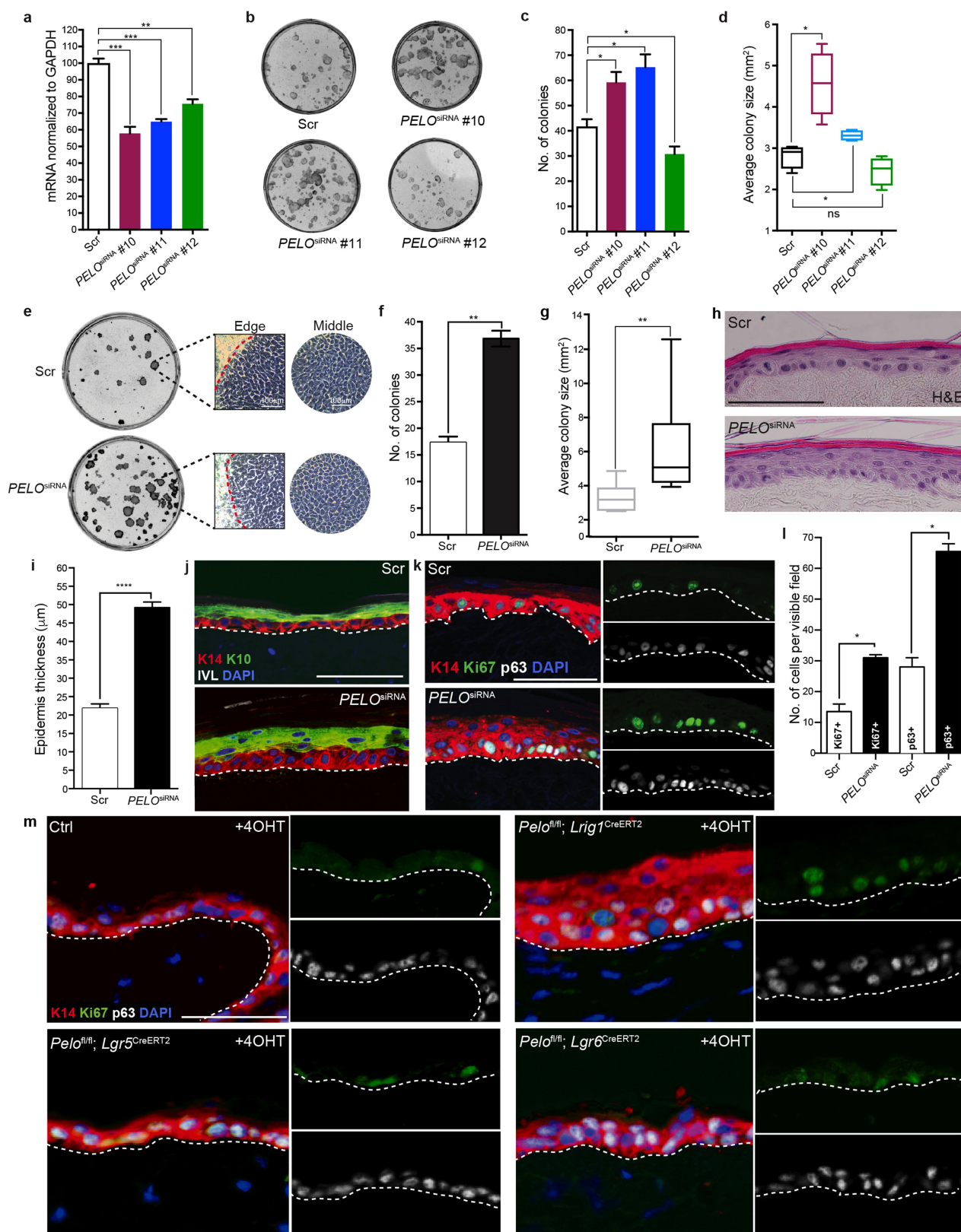
Extended Data Fig. 3 | *Pelo* deletion leads to progressive hair follicle and sebaceous gland abnormalities. a–c, Confocal images of tail epidermal wholemounts immunostained for Krt14, hair follicle bulge markers CD34 and Krt15, sebocyte maturation marker fatty acid synthase (FASN) and junctional zone stem cell marker Lrig1 show progressive

changes in hair follicle and sebaceous gland structure from P16 to P120 in *Pelo*^{epiKO} mice. Note that the FASN staining in P84 and P120 *Pelo*^{epiKO} epidermis is non-specific owing to highly keratinized hair follicles. Asterisks in **b** indicate non-specific staining of sebaceous glands. Scale bars, 100 μ m.



Extended Data Fig. 4 | Postnatal epidermal *Pelota* deletion impairs barrier function and wound healing. **a**, **b**, Breeding scheme and topical 4-OHT treatment regimen. **c**, Representative *Pelota^{fl/fl}; K14^{CreERT}* mouse showing skin lesions (dashed area) in 4-OHT-treated dorsal skin. **d**, TEWL is increased in 4-OHT-treated skin of *Pelota^{fl/fl}; K14^{CreERT}* mice. **e**, Rate of wound closure. **f**, Tail epidermal wholemounts immunostained with Krt14 and Krt15 antibodies showing altered sebaceous gland architecture (arrows) in 4-OHT-treated *Pelota^{fl/fl}; K14^{CreERT}* mice. **g**, Tail epidermal wholemounts from TdTomato (red) genetically labelled *Pelota^{fl/fl}; K14^{CreERT}*

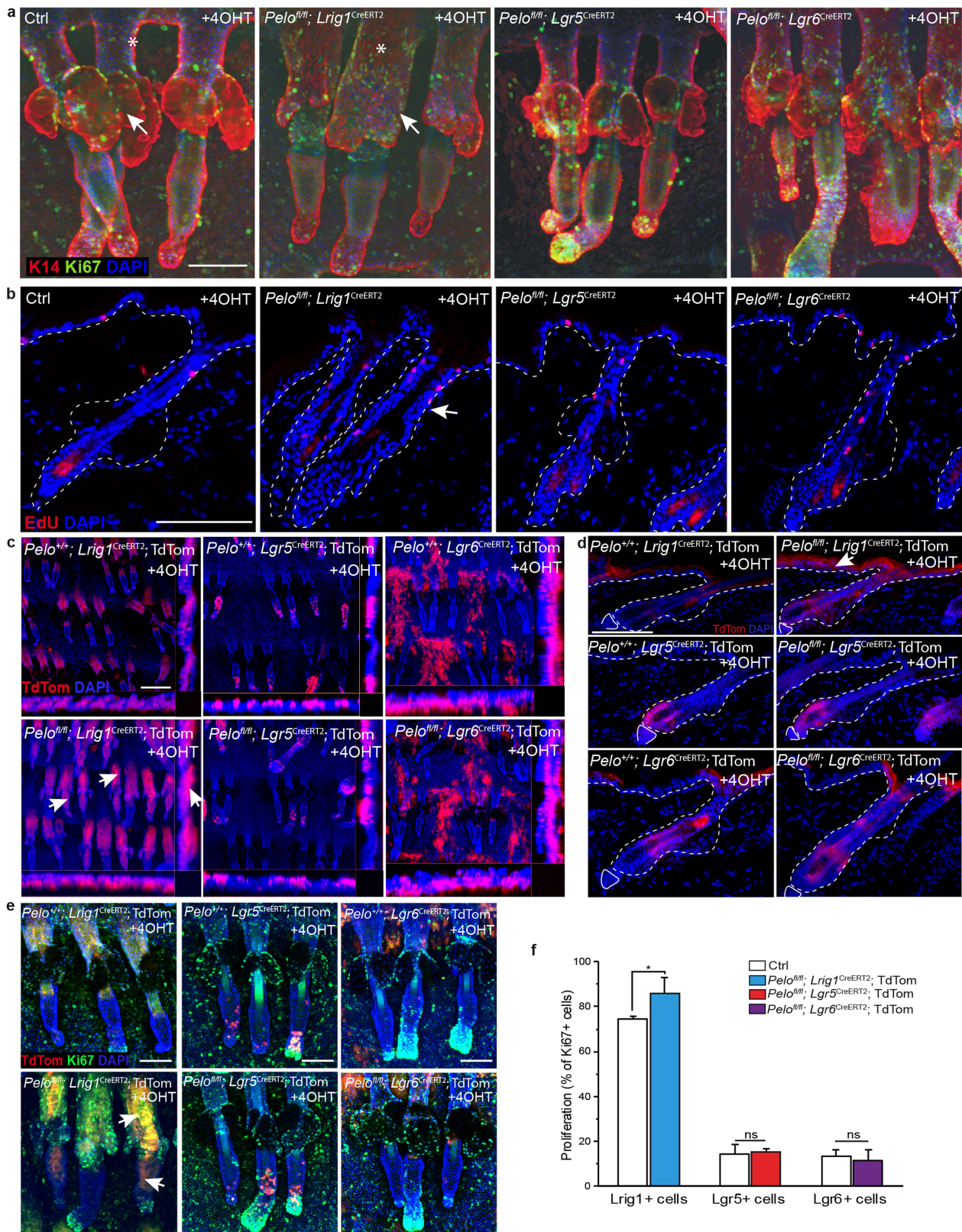
mice show keratinized cysts in hair follicles (arrows). **h**, Cumulative mean values of gene expression obtained from ribosome profiling show downregulation of markers of sebaceous gland differentiation and increase in *Myc*. **i**, Tail epidermal wholemounts showing altered expression of FASN, Scd1 and Lrig1 (arrows) in sebaceous glands of 4-OHT-treated *Pelota^{fl/fl}; K14^{CreERT}* mice (middle and right). Dashed lines indicate pilosebaceous units. Scale bars, 100 μ m. ** $P = 0.0072$, * $P = 0.0650$, ns, not significant. $n = 3$ in treated and untreated control groups.



Extended Data Fig. 5 | See next page for caption.

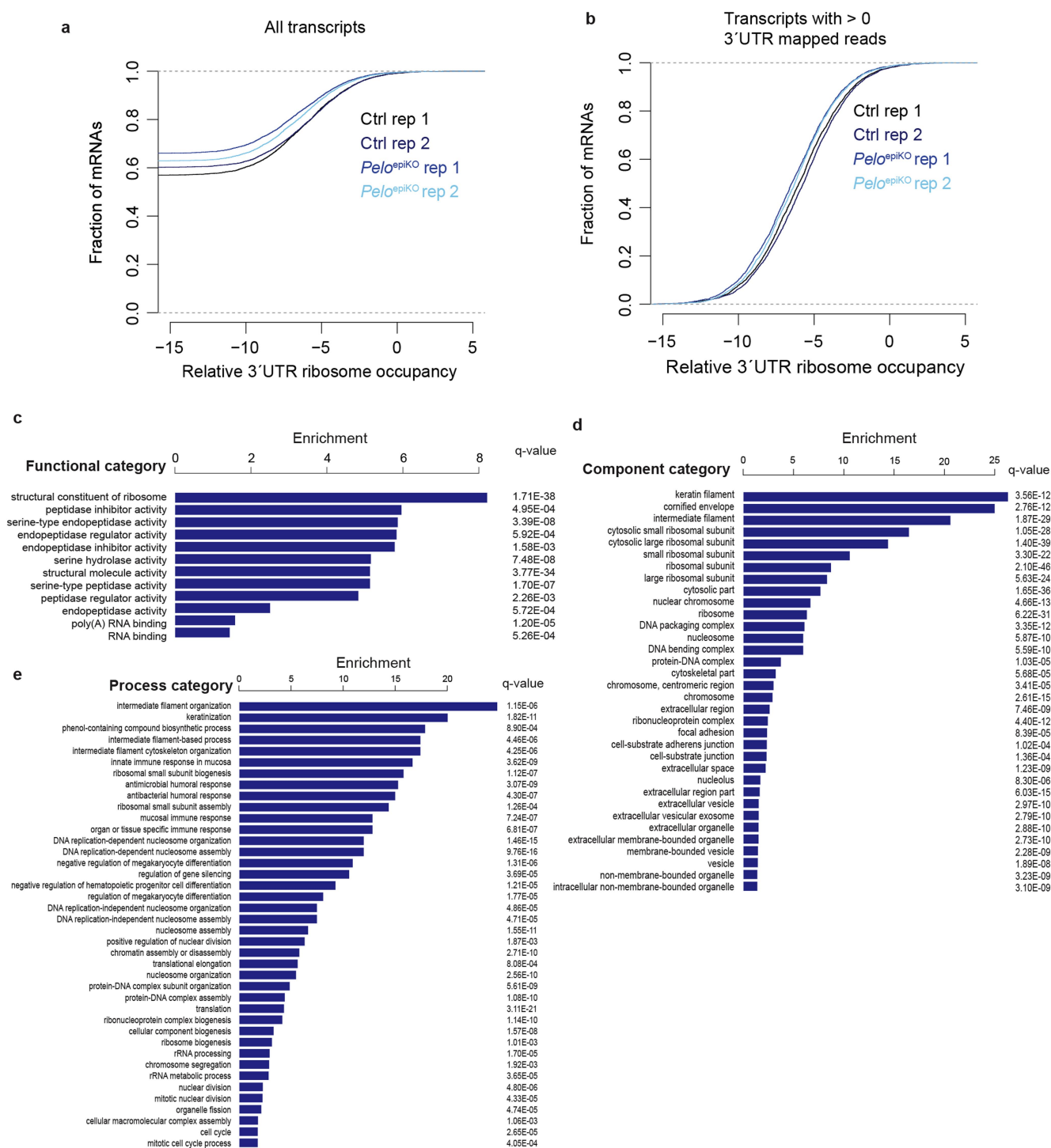
Extended Data Fig. 5 | Knockdown of *PELO* in human keratinocytes phenocopies mouse epidermal phenotype and proliferation difference in mice lacking *Pelo* in *Lrig1*⁺, *Lgr5*⁺ and *Lgr6*⁺ stem cells. a–d, *PELO* knockdown validation. **a**, qRT–PCR for individual siRNAs transfected in human primary keratinocytes. **b**, Clonal growth. **c**, **d**, Colony number and average size of individual colonies. **e–g**, Clonal growth of keratinocytes, comparing pooled *PELO* siRNA knockdown (*PELO*^{siRNA}) and scrambled (Scr) control. **h–l**, Effect of *PELO* knockdown in human epidermal reconstitution assay on decellularized dermis. **h**, **i**, Epidermal thickness of de-epidermized dermis (DED) cultures is significantly increased on *PELO* knockdown. **j–l**, Immunolabelling for Krt14 (K14), Ki67, p63

and differentiation markers Krt10 (K10) and involucrin (IVL) shows increased number of differentiated cell layers (**j**) and increased number of cells expressing Ki67 and p63 (**k**, **l**) in *PELO*^{siRNA} reconstituted epidermis. Dashed lines indicated dermal–epidermal boundary. **m**, Assessing proliferation by Ki67 and p63 labelling in dorsal IFE sections of mice lacking *Pelo* in *Lrig1*, *Lgr5* and *Lgr6* stem cells. Scale bars, 100 μ m. ****P* = 0.0009 (**a**, for siRNA#10), ****P* = 0.0004 (**a**, for siRNA#11), ***P* = 0.0031 (**a**, for siRNA#12); **P* = 0.0286 (**c**); **P* = 0.0286 (**d**); ***P* = 0.0022 (**f**); ***P* = 0.0087 (**g**); *****P* \leq 0.0001 (**i**); **P* = 0.0229 for Ki67 and **P* = 0.0107 for p63 (**l**). *n* = 2 independent transfections; *n* = 3 dishes (**a–g**) and *n* = 2 sections of reconstituted epidermis (**h**, **l**).



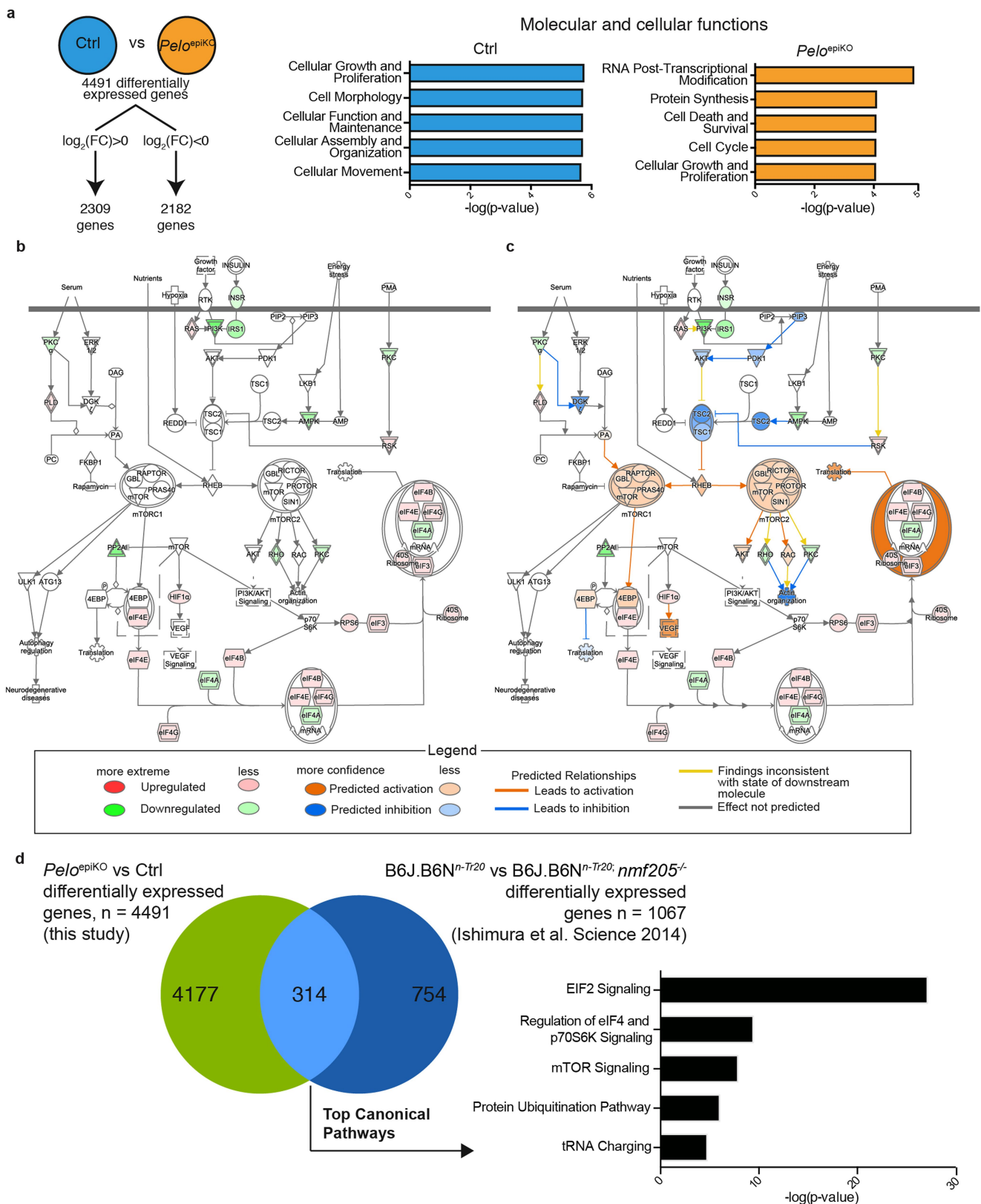
Extended Data Fig. 6 | *Lgr1*⁺ stem cells account for *Pelo* mutant epidermal phenotype. **a**, Tail epidermal wholemounts labelled with Krt14 and Ki67 antibodies, showing increased proliferation and alterations to the junctional zone (asterisks) and sebaceous glands (arrow) in *Pelo^{fl/fl} Lgr1^{CreERT2}* mice. **b**, Cross-section of dorsal skin stained for EdU shows increased proliferation and alterations in hair follicle infundibulum structure (arrow) in *Pelo^{fl/fl} Lgr1^{CreERT2}* mice. **c–e**, Confocal images of tail epidermal wholemounts (**c**, **e**) and dorsal skin sections (**d**) of tdTomato

labelled *Pelo^{fl/fl} Lgr1^{CreERT2}*, *Pelo^{fl/fl} Lgr5^{CreERT2}* and *Pelo^{fl/fl} Lgr6^{CreERT2}* mice. **c**, **d**, Expansion of tdTomato-labelled *Lgr1* (arrows) but not *Lgr5* or *Lgr6* progeny upon *Pelo* deletion. **e**, **f**, Increase in proliferation (Ki67 labelling) of *Lgr1*⁺ (arrows) but not *Lgr5*⁺ and *Lgr6*⁺ populations. Scale bars, 100 μ m. **P* = 0.0047 (**f**). *n* = 9 wholemounts analysed over 3 mice per group. All mice were in telogen of the hair cycle (2–3 months old) when treated with 4-OHT. Treatment regime and harvest of tissue were as indicated in Fig. 2c. Dashed lines mark epidermal–dermal boundary.



Extended Data Fig. 7 | *Pelo* knockout epidermal cells do not accumulate 3' UTR footprints. a, b, Empiric cumulative distribution plots of relative 3' UTR ribosome occupancy for all transcripts (a) or those with at

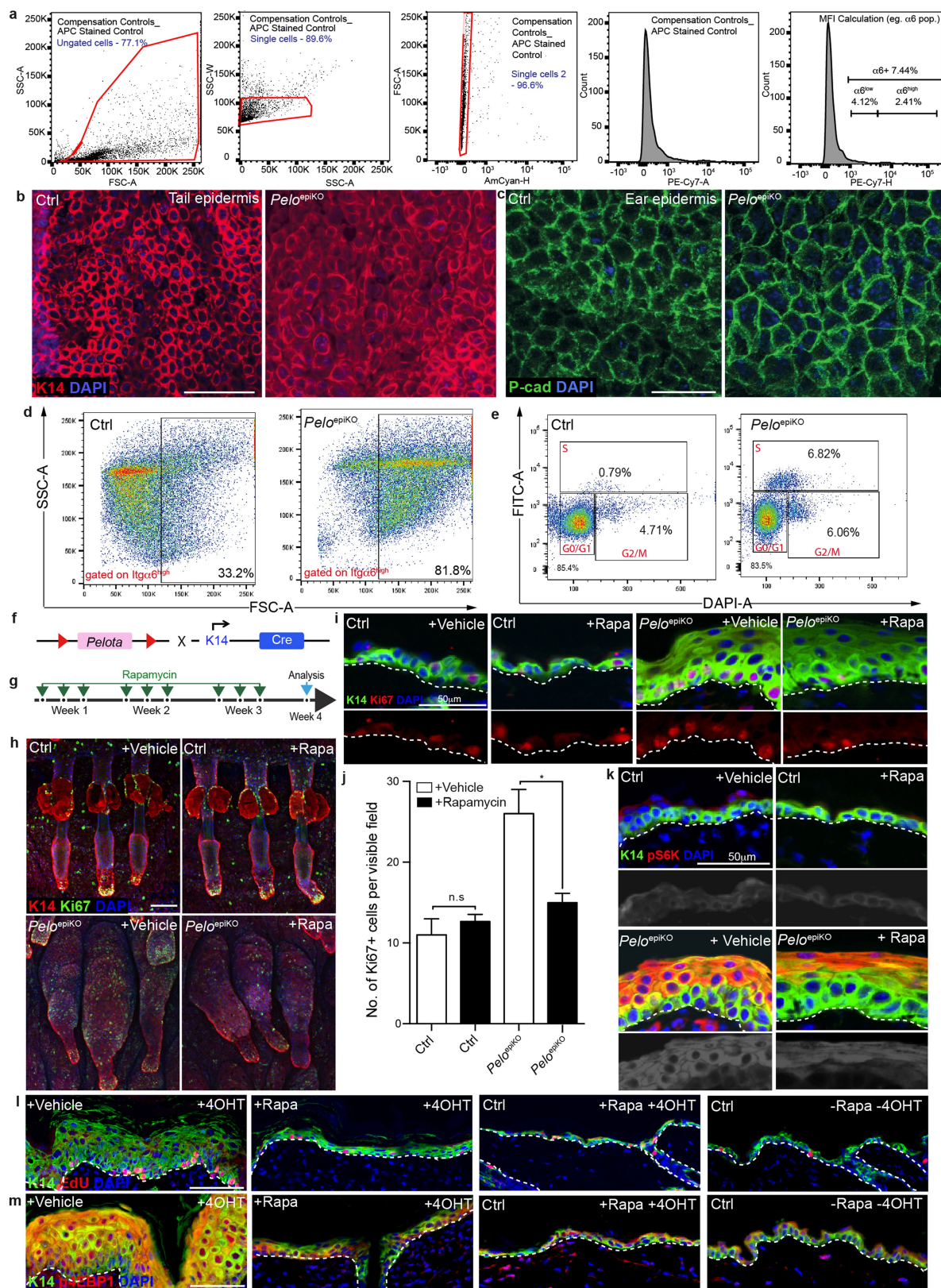
least one read mapped to the 3' UTR (b). c–e, Gene Ontology of genes differentially expressed in *Pelo*-null epidermis. Functional, component and process categories of genes enriched in *Pelo*^{epiKO} mice.



Extended Data Fig. 8 | Computational analysis of differentially regulated pathways between control and *Pello*^{epiKO} and comparison of molecular signatures in *Pello*^{epiKO} and *Gtbp2*-deficient brain.

a, Number of genes that were differentially expressed in *Pello*^{epiKO} and control epidermis and their associated functions. **b, c**, Ingenuity Pathway Analysis showing changes in mTOR pathway genes in *Pello*^{epiKO} versus

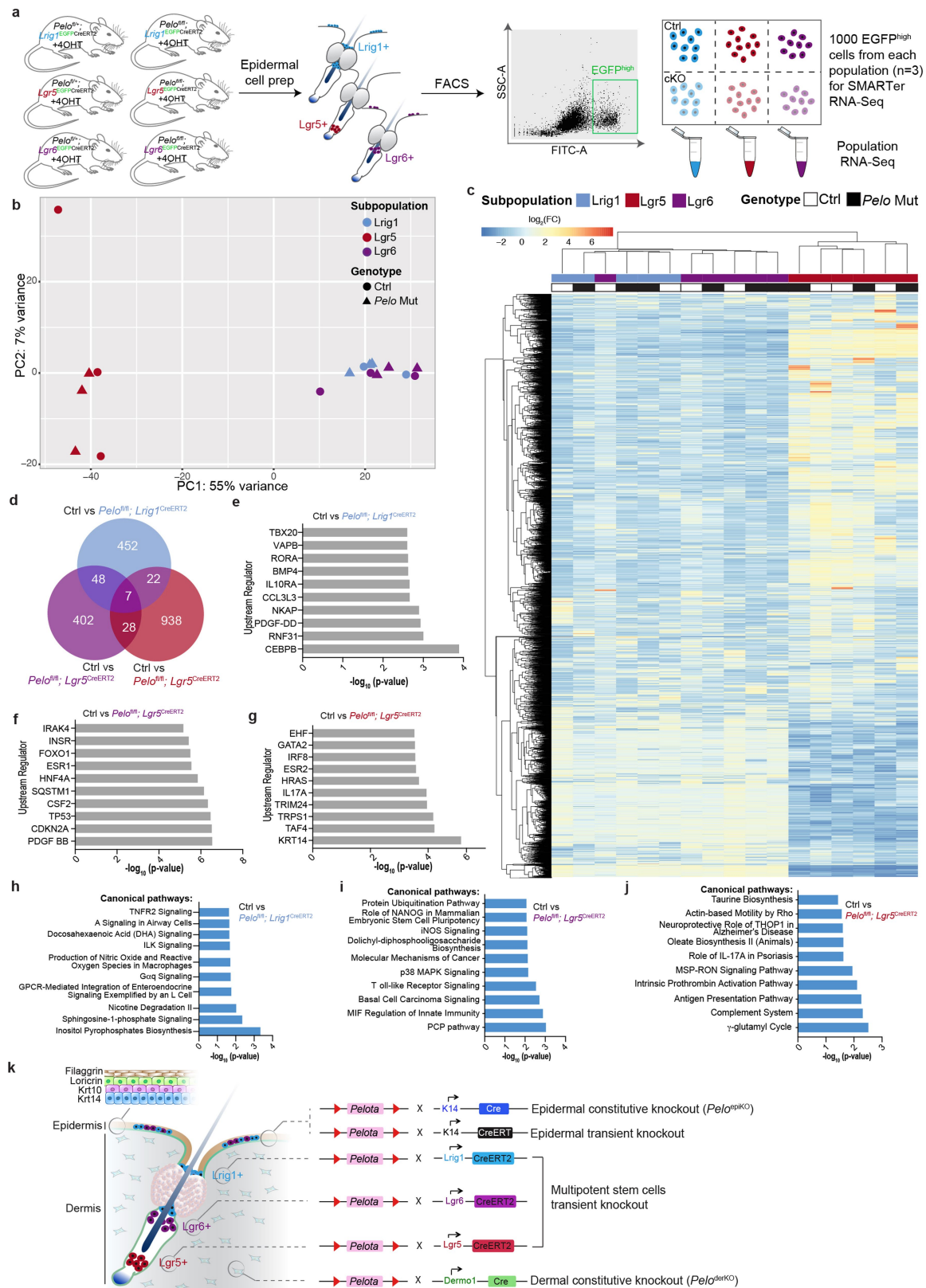
control epidermis (**b**) and their predicted molecular activities (**c**). **d**, Venn diagram shows common differentially expressed genes in our study and that of Ishimura et al.⁸ when comparing control and mutant mice. The 314 overlapping genes are enriched in top canonical pathways that are highly related to translation.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | *Pelo* epidermal deletion results in increased protein synthesis and basal stem cell size and rapamycin treatment reduces proliferation of *Pelo*-null epidermis. **a**, Gating strategy for measurement of OP-P incorporation into cell populations. **b, c**, Confocal images of tail and ear epidermal wholemounts immunolabelled for Krt14 and P-cadherin (P-Cad), showing IFE basal cells. **d, e**, Representative flow cytometric dot plot showing increased cell size (FSC-A) of Itg α 6^{high} cells and altered S and G2/M cell cycle phases in *Pelo*^{epiKO} epidermis. **f, g**, Breeding scheme and rapamycin treatment regimen. **h**, Immunolabelling of tail epidermal wholemounts with Krt14 and Ki67 antibodies shows reduced proliferation in rapamycin-treated mice compared to vehicle-treated group. Note that there was no significant change in epidermal proliferation of control mice treated with rapamycin when compared to vehicle-treated mice. **i, j**, Cross-sections of IFE from *Pelo*^{epiKO} and control back skin immunolabelled with Krt14 and Ki67 antibodies, showing significant reduction in Ki67⁺ and suprabasal Krt14⁺ cells in rapamycin-

treated compared to vehicle-treated mice. **k**, Cross-sections of IFE from control and *Pelo*^{epiKO} back skin immunolabelled with Krt14 and pS6K antibodies showing marked increase in pS6K labelling, indicating mTOR hyperactivation in vehicle-treated *Pelo*^{epiKO} skin. **l**, Cross-sections of IFE of control and *Pelo*^{fl/fl}Krt14^{CreERT} mice (with simultaneous 4-OHT and rapamycin treatment) immunolabelled for Krt14 and EdU showing significant reduction in EdU⁺ and suprabasal Krt14⁺ cells in rapamycin-treated compared to vehicle-treated mice. **m**, Cross-sections of IFE of control and *Pelo*^{fl/fl}Krt14^{CreERT} mice (with simultaneous 4-OHT and rapamycin treatment) immunolabelled for Krt14 and p4EBP antibodies. Note reduced pS6K labelling (**k**) and p4EBP1 (**m**) in rapamycin-treated epidermis. Greyscale images for pS6K are shown below merged images. Scale bars, 100 μ m. * P = 0.0132 (**j**), n. s., non significant. n = 9 sections analysed over 3 mice per group. Dashed lines mark epidermal–dermal boundary.

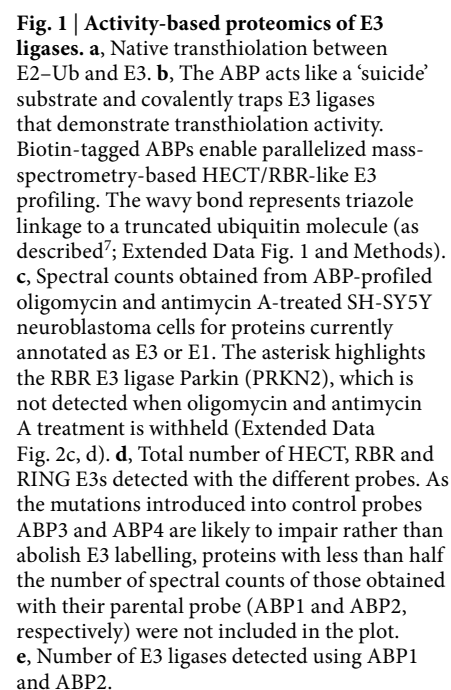


Extended Data Fig. 10 | RNA-seq of Lgr1⁺, Lgr5⁺ and Lgr6⁺ cells reveals Lgr5 as a transcriptionally unique subpopulation and subtle changes in transcription in all subpopulations when *Pelo* is deleted. **a**, Schematic illustration of the EGFP^{high} sorting and RNA-seq strategy for control and *Pelo*-deleted subpopulations using *Pelo*^{fl/fl}; Lgr1^{EGFP-CreERT2}, *Pelo*^{fl/fl}; Lgr5^{EGFP-CreERT2} and *Pelo*^{fl/fl}; Lgr6^{EGFP-CreERT2} mice. **b**, Principal component analysis of RNA-seq data shows that the Lgr5 subpopulation is remarkably different from the other two. Note that there is no major change in the clusters when *Pelo* is deleted in any of the subpopulations. **c**, Hierarchical clustering of the subpopulations corroborates minimal transcriptional changes between control and Mut mice, revealing two

major clusters, one for Lgr5 and another for Lgr1 and Lgr6. **d**, Venn diagram illustrating the differentially expressed genes in common among the three subpopulations when comparing control and mutant cells. **e–g**, Top differentially regulated transcription factors between control and *Pelo*^{fl/fl}; Lgr1^{EGFP-CreERT2} (**e**), *Pelo*^{fl/fl}; Lgr5^{EGFP-CreERT2} (**f**) and *Pelo*^{fl/fl}; Lgr6^{EGFP-CreERT2} (**g**) subpopulations. **h–j**, Top differentially regulated canonical pathways between control and *Pelo*^{fl/fl}; Lgr1^{EGFP-CreERT2} (**h**), *Pelo*^{fl/fl}; Lgr5^{EGFP-CreERT2} (**i**) and *Pelo*^{fl/fl}; Lgr6^{EGFP-CreERT2} (**j**) subpopulations. **k**, Schematic of skin showing location of marker expression and the various transgenic mice used in this study.

Kuan-Chuan Pao¹, Nicola T. Wood¹, Axel Knebel¹, Karim Rafie², Mathew Stanley², Peter D. Mabbitt¹, Ramasubramanian Sundaramoorthy², Kay Hofmann³, Daan M. F. van Aalten² & Satpal Virdee^{1*}

We prepared biotinylated variants of activity-based probes (ABPs)⁶, which profile the hallmark transthioleation activity of HECT/RBR E3 ligases⁷ (Fig. 1a, Extended Data Fig. 1). Combining the ABP technology with mass spectrometry enabled parallelized profiling of E3 ligase activity in neuroblastoma SH-SY5Y cell extracts^{8,9} (Fig. 1b and Supplementary Information). E3 ligases were filtered using criteria to ensure that signals for at least a subset of detected E3 ligases correlated with E3 ligase activity and/or abundance (Extended Data Fig. 2). We successfully profiled around 80% of the approximately 50 known



¹MRC Protein Phosphorylation and Ubiquitylation Unit, University of Dundee, Dundee, UK. ²Division of Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, UK. ³Institute for Genetics, University of Cologne, Cologne, Germany. *e-mail: s.s.virdee@dundee.ac.uk

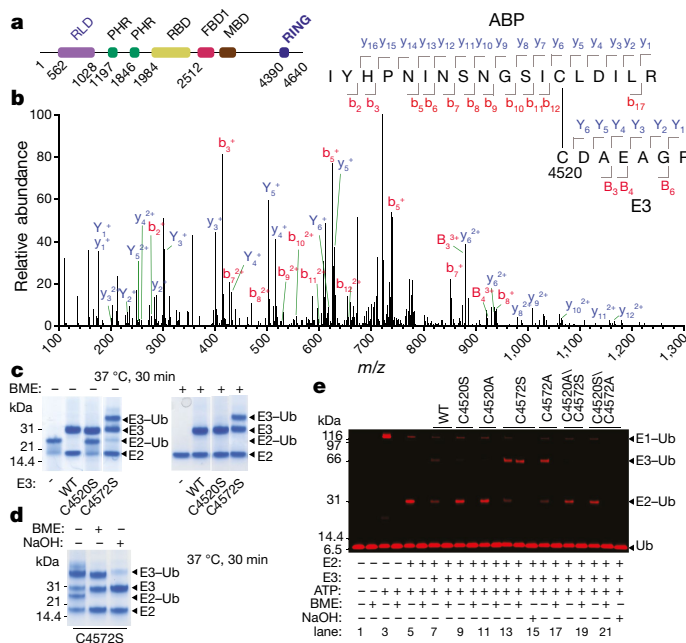


Fig. 2 | MYCBP2 is a novel class of E3 ligase and data support a cysteine relay mechanism. **a**, Domain architecture of MYCBP2: RCC1-like GEF domain (RLD), two PHR-family-specific (PHR) domains, a RAE1 binding domain (RBD), an F-box binding domain 1 (FBD1), a Myc binding domain (MBD) and a C-terminal RING domain¹⁰. **b**, Representative tandem mass spectrum (MS/MS spectrum) for a crosslinked peptide derived from ABP-treated wild-type MYCBP2(cat), showing that C4520 is labelled. The spectrum is for a 5^+ precursor ion. Observed $m/z = 614.5088$; theoretical $m/z = 614.5094$. Top right, the E2-derived peptide from the ABP and the peptide from MYCBP2(cat). The solid line corresponds to the crosslink sites and an additional ABP remnant⁷. **c**, Multiple-turnover E2 discharge assay onto hydroxy nucleophiles (Tris and glycerol) present in the reaction buffer. BME, 2-mercaptoethanol in gel loading buffer; WT, wild-type. **d**, The ubiquitin adduct on MYCBP2(C4572S) is base-labile (0.14 N NaOH), indicative of the formation of an engineered (oxy)ester linkage between ubiquitin and S4572 on MYCBP2. **e**, SDS-PAGE thioester-ester trapping assay, tracking Cy3B-labelled ubiquitin. All biochemical data were consistent across three biological replicates with the exception of the mass spectrometry experiment which was carried out once.

HECT/RBR E3 ligases by this method, but unexpectedly, 33 RING E3 ligases, which lack HECT or RBR ancillary domains, were also enriched (Fig. 1c–e). To explore the possibility that previously undiscovered RING-linked E3 ligases were being labelled, we focused on MYCBP2 (Myc-binding protein 2), also known as PHR1 (PAM/Highwire/Rpm-1) (Fig. 1c). MYCBP2 is a large, 0.5-MDa neuron-associated protein, which contains a C-terminal RING domain (Fig. 2a) and is involved in a range of cellular processes including regulation of nervous system development and axon degeneration^{10–12}.

MYCBP2(cat), a recombinant C-terminal version of MYCBP2 encompassing the RING domain (residues 4378–4640; Fig. 2a) and an uncharacterized C-terminal cysteine-rich region underwent robust ABP labelling with an efficiency comparable to that of E3 ligases known to demonstrate transthiolation activity^{7,13} (Extended Data Fig. 3a). To map the putative catalytic cysteine, we used a combination of ABP-based profiling and ABP-crosslinking mass spectrometry⁷ (Fig. 2b, Extended Data Fig. 3b, c; full gels and blots in Supplementary Fig. 1). The results of these experiments supported the hypothesis that C4520 is a putative catalytic residue. We next assayed wild-type E3 ligase activity, but were unable to detect autoubiquitination or free ubiquitin chain formation. However, we observed rapid E3-dependent discharge of ubiquitin from E2–Ub, suggesting the presence of an unknown small-molecule nucleophilic acceptor (Fig. 2c). Liquid chromatography–mass spectrometry (LC–MS) analysis revealed that ubiquitin was being quantitatively converted into two species with masses corresponding

to condensation products with tris(hydroxymethyl)aminomethane (Tris) and glycerol (8,639 Da and 8,668 Da, respectively), both of which were present in our assay buffer at routinely employed concentrations of 50 mM and around 65 mM, respectively. Owing to the common hydroxy functionality within these nucleophiles, MYCBP2 appeared to show esterification activity (Extended Data Fig. 4a, b); this activity was found to be dependent on C4520, consistent with it forming a thioester-linked E3–Ub intermediate^{4,5} (Fig. 2c).

Unexpectedly, a MYCBP2(C4572S) mutant retained activity while also forming a discrete mono-ubiquitin adduct that was resistant to thiolysis but was reversible after base treatment⁵ (Fig. 2c, d, Extended Data Fig. 4c). One possible explanation for this was that the mutant S4572 residue contributed to catalysis by forming an additional and less transient (oxy)ester-linked intermediate between ubiquitin and S4572 that retained the ability to modify substrate. We hypothesized that C4520 and C4572 are both catalytic residues that function in tandem by relaying ubiquitin from one cysteine to the other through an intramolecular transthiolation reaction. To test this relay mechanism, we carried out gel-based thioester-ester trapping assays¹⁴ (Fig. 2e, Extended Data Fig. 4d) and observed a thiol-sensitive ubiquitin adduct on wild-type MYCBP2(cat), which was not observed with the C4520S mutant (Fig. 2e). Consistent with the earlier experiments (Fig. 2c), C4572S underwent adduct formation that was thiol-resistant but base-labile (Fig. 2e). Thus, if a transient thioester intermediate was being formed between ubiquitin and C4520, an unreactive C4572A mutant should stabilize it. Indeed, thiol-sensitive adduct formation was increased on a C4572A mutant relative to the wild type and was presumably linked via a thioester bond (Fig. 2e). Adduct formation was not detected with a C4520A/C4572S double mutant, consistent with ubiquitin being linked to the C4520 residue (Fig. 2e). In the absence of a direct demonstration of Cys-to-Cys ubiquitin transfer, we cannot formally exclude other possibilities. However, the existing data are consistent with the essential cysteines functioning in a relay mechanism. Mutational analysis and size exclusion chromatography–multi-angle light scattering (SEC–MALS) of MYCBP2(cat) (Extended Data Fig. 4e, f) suggests that the proposed relay mechanism requires that both essential cysteines are in the same molecule, consistent with an intramolecular-cis-relay mechanism.

In light of the observed esterification activity, we attempted to identify the amino acid substrate of MYCBP2 by screening a panel of amino acids⁵, and found that discharge activity was markedly enhanced with threonine. Product formation was dependent on C4520 (Fig. 3a, b, Extended Data Fig. 5a–d). Mass-spectrometry-based quantification indicated approximately tenfold enhanced selectivity for threonine over serine (Extended Data Fig. 5e). Although we observed a low level of lysine modification, this was independent of MYCBP2(cat)⁵. A threefold selectivity for threonine was also maintained in a peptide context (Fig. 3c, Extended Data Fig. 5f–h). Furthermore, basal ubiquitination of a lysine peptide was partially inhibited in the presence of MYCBP2(cat), underscoring its lack of activity towards lysine (Fig. 3c). Taken together, our experiments revealed that MYCBP2 is a member of a novel class of E3 ligase that operates via two essential cysteines, promotes ubiquitin modification of hydroxyl groups, and preferentially esterifies threonine over serine with ubiquitin. As MYCBP2 uses a novel mechanism, we termed it a RING-Cys-relay (RCR) E3 ligase. We measured the catalytic efficiency of the MYCBP2-threonine-esterification activity and found that it had an intermediate value between those of well-characterized HECT (UBE3C) and RBR (HHARI) E3 lysine aminolysis activity^{5,15} (Extended Data Fig. 5i–k). E2 mutational analysis^{5,16–19} provided further support for MYCBP2(cat) being a member of a novel class of E3 ligase (Extended Data Fig. 6a, b and Methods). We tested 17 E2 conjugating enzymes to identify functional E2 partners, but only UBE2D1, UBE2D3 and UBE2E1 exhibited robust activity (Extended Data Fig. 6c).

MYCBP2 promotes Wallerian axon degeneration through destabilization of nicotinamide mononucleotide adenylyltransferase (NMNAT2)²⁰. We next tested whether MYCBP2(cat) can ubiquitinate NMNAT2 by esterification in vitro (Fig. 3d). Despite containing 13 lysine residues, NMNAT2 underwent hydroxide-labile but

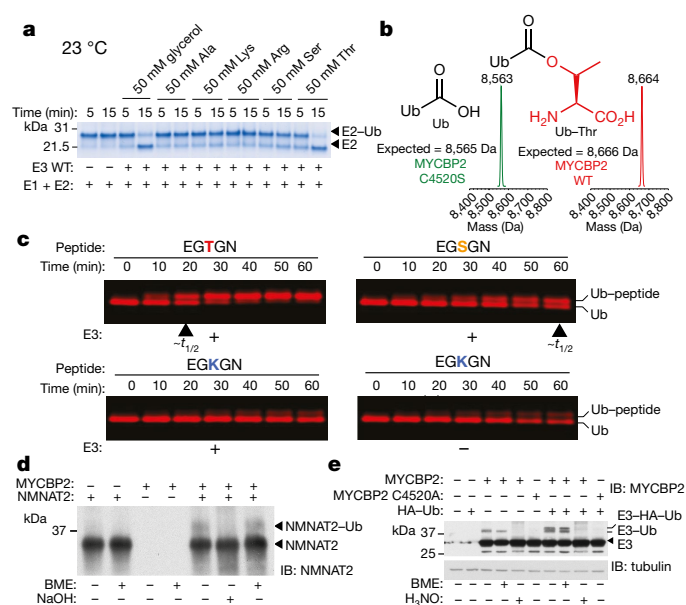


Fig. 3 | MYCBP2 ubiquitinates serine and threonine with selectivity for threonine. **a**, E3-mediated multiple-turnover discharge reaction for a panel of amino acids (50 mM). Similar results were obtained over three independent experiments. **b**, Deconvoluted mass spectrum of ubiquitin species after reaction with threonine. Observed molecular weight of ubiquitinated threonine (Ub–Thr) = 8,664 Da; theoretical molecular weight = 8,666 Da. Observed molecular weight of unmodified ubiquitin = 8,563 Da; theoretical molecular weight = 8,565 Da. Note that Ub–Thr can undergo O–N acyl transfer, forming the peptide-linked species. **c**, SDS–PAGE analysis of peptide modification with fluorescently labelled ubiquitin. Based on half-life approximation, MYCBP2(cat) has a threefold selectivity for threonine over serine in the depicted peptide context. Similar results were obtained in two independent experiments. **d**, Recombinant NMNAT2 undergoes base-labile ubiquitination in a MYCBP2(cat)-dependent manner. Similar results were obtained in two independent experiments. **e**, MYCBP2(cat) transiently transfected into HEK293 cells undergoes hydroxy-ubiquitination that is dependent on C4520. H₃NO, treatment with 0.5 M hydroxylamine at pH 9.0; IB, immunoblot. Similar results were obtained over three biological replicates.

thiol-resistant ubiquitination, demonstrating that MYCBP2 can target hydroxy residues within one of its putative substrates²⁰. Cellular substrate recognition is mediated by a Skp1–Fbxo45 substrate receptor co-complex that binds to a site approximately 1,940 residues N-terminal to the MYCBP2(cat) region²¹ (Fig. 2a). NMNAT2 also undergoes palmitoylation and rapid axonal transport²², making reconstitution and cellular study of its ubiquitination extremely challenging. However,

to establish whether MYCBP2(cat) retains non-lysine activity in cells, we investigated autoubiquitination of MYCBP2(cat) after transient transfection into HEK 293 cells. Base-labile (but thiol-resistant) ubiquitination was observed that was dependent on C4520 (Fig. 3e). This demonstrates that MYCBP2 can retain specificity for hydroxy amino acids in cells and that this activity remains dependent on the upstream catalytic residue that is implicated in the ubiquitin relay mechanism.

To further validate the RCR model and the serine and threonine activity, we crystallized MYCBP2(cat) (residues 4378–4640) and solved the crystal structure to a resolution of 1.75 Å (Extended Data Table 1 and Extended Data Fig. 7a–c). Residues 4388–4441 at the N terminus correspond to the predicted cross-brace C3H2C3 RING domain (Fig. 4a, Extended Data Fig. 7d). Following the RING domain is a long α -helix (4447–4474) that leads into a small helix–turn–helix motif (residues 4475–4500) (Fig. 4a, Extended Data Fig. 7e), and further C-terminal is a globular domain (residues 4501–4638) that binds four Zn ions (Fig. 4b, Extended Data Fig. 7f, g). Since this domain also contains the two essential catalytic residues, we designate it the tandem cysteine domain. Between the β A2 strand and helix 3₁₀A is an unstructured region (4519–4526) that projects out to the side of the core Zn-binding fold. The upstream C4520 residue resides within this unstructured region, which, together with flanking residues, forms a mobile region that we term the mediator loop. The configuration of Zn coordination (C5HC7HC2) in the tandem cysteine domain is semi-contiguous and does not adopt cross-brace architecture (Extended Data Fig. 7f).

Crystal packing revealed that T4380, within the N terminus of a symmetry-related MYCBP2(cat) molecule (T4380(sym)), is located proximal to the esterification site where it forms a number of substrate-like interactions (Fig. 5a, b). First, the β -hydroxy group of T4380(sym) complements E4534 and H4583 and forms a potential triad (Fig. 5a). Thus the β -oxygen atom of T4380(sym) appears to be primed for deprotonation and nucleophilic attack. The C terminus of ubiquitin, when thioester-linked to C4572, is a catalytically productive electrophilic centre. Even though this ubiquitin molecule is absent from our structure, the sulfur atom in C4572 is 3.8 Å away from the β -oxygen atom of T4380(sym). Therefore, the structure appears to accurately reflect a catalytic intermediate that is poised to undergo threonine ubiquitination by esterification of its β -hydroxy group (Fig. 5a). Furthermore, a sub-cluster of phenylalanine residues (F4573, F4578 and F4586), proximal to the β -methyl group of T4380(sym) (Fig. 5a, b), forms a well-defined hydrophobic pocket that the β -methyl group of T4380(sym) docks into, and seems to be a positive determinant of selectivity for the threonine side chain. The proposed roles of these residues were validated in threonine discharge assays (Fig. 5c). A H4583N mutation abolished activity consistent with a role for this residue as a general base. The H4583N mutant also underwent enhanced, thiol-sensitive

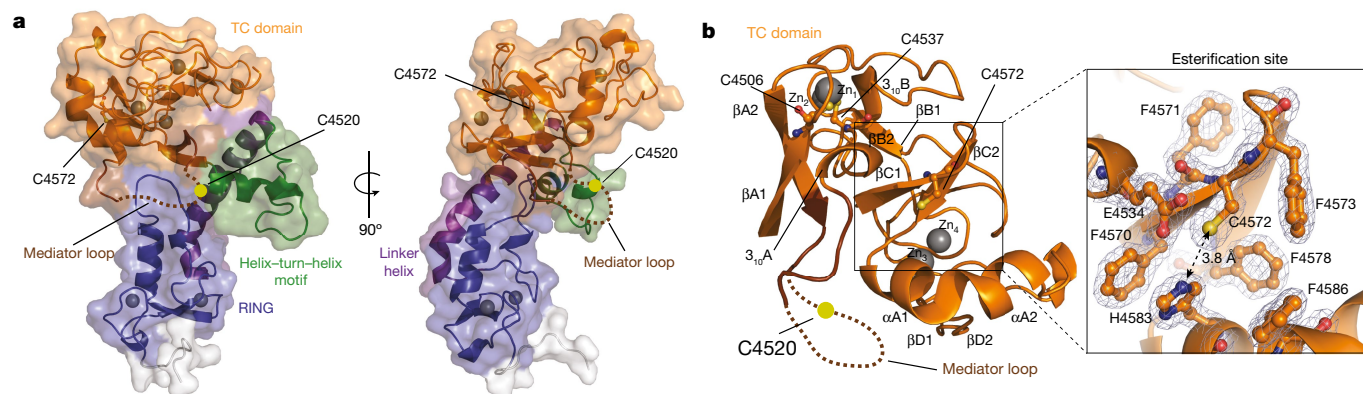


Fig. 4 | Crystal structure of MYCBP2(cat). **a**, Surface and cartoon representation of the structure refined to 1.75 Å. **b**, The tandem cysteine (TC) domain and its secondary structure annotation. Inset shows a close-up of the esterification site, in which a triad-like arrangement

centres on C4572. A $2|F_{\text{obs}}| - |F_{\text{calc}}|$ electron density map contoured at 1.5 σ is represented in mesh for esterification-site residues. C4506 and C4537, which abolish ABP labelling when mutated, are Zn-coordinating residues.

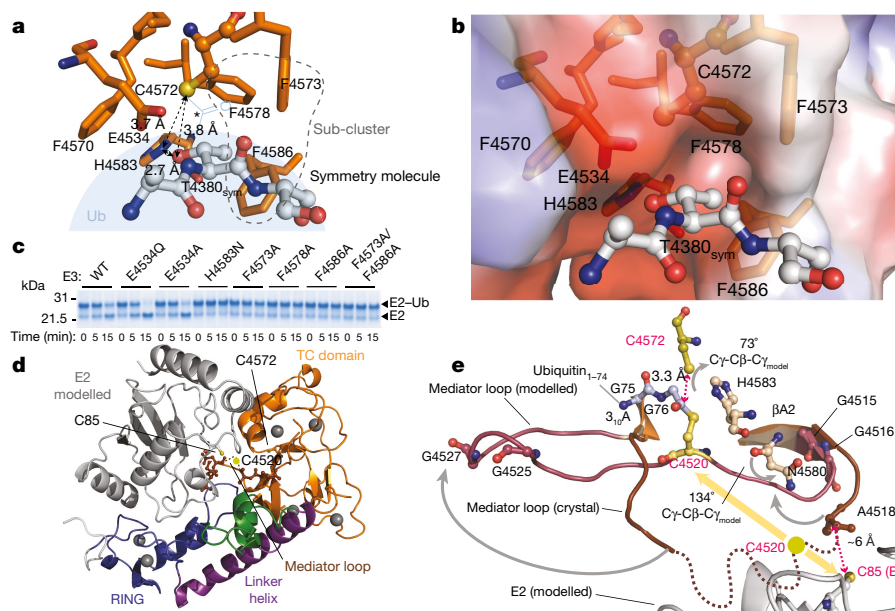


Fig. 5 | Structural basis for threonine selectivity, model of an E2–E3 intermediate and model of ubiquitin relay. **a**, A threonine residue at the N terminus of a symmetry-related molecule (T4380(sym)) in an Ala–Thr–Ser sequence motif (grey ball and stick representation) is docked into the esterification site (orange sticks). A ubiquitin thioester-linked to C4572 is shown in light blue. The asterisk shows the position of the electrophilic centre of the ubiquitin thioester carbonyl–carbon atom. **b**, Electrostatic potential of the esterification site. Blue, positive potential; red, negative potential; grey, neutral. **c**, Residues hypothesized to be important for catalysis were mutated and tested in threonine discharge assays. **d**, Superposition of the RING domain from the RBR E3 ligase

ubiquitin adduct formation in accordance with the anticipated defect in rendering substrate nucleophiles reactive towards the C4572 thioester (Extended Data Fig. 8a). Conservative perturbation of the phenylalanine cluster also markedly reduced threonine discharge activity (Fig. 5c). However, perturbation of E4534 did not reduce activity; consequently, its precise role remains unclear.

Conservation of RING-domain binding to E2^{16–18,23} permitted the modelling of an E2–RCR E3 ligase complex that was geometrically compatible with transthiolation between E2–Ub and C4520 in MYCBP2(cat) (Fig. 5d, Extended Data Fig. 8b). To simulate the conformation required for subsequent ubiquitin relay to C4572, we modelled the missing mediator-loop residues with a Gly–Gly dipeptide thioester linked to C4520, representative of the C terminus of ubiquitin that would be transferred via transthiolation with E2–Ub if our relay model was valid (Extended Data Fig. 8c–e). Consistent with this mechanism, the carbonyl–carbon of the ubiquitin thioester could be positioned in the proximity (around 3 Å) of the C4572 sulfhydryl sulfur atom. To adopt this conformation, minor twisting of a Gly–Gly motif (residues 4515–4516) at the tip of the β A2 strand was necessary. Clashes were observed between mediator-loop residues further C-terminal and R4533, E4534, N4580, H4583 and D4584, but these could largely be relieved by rotations of their side chains into available space (Extended Data Fig. 8d). As ordered loop residues 4527–4531 required a substantial displacement to generate the model, we speculate that the mobile mediator loop region would span residues 4515–4531. As C4520, which resides within this mobile structural element, needs to be engaged by the E2 active site²⁴ this might account for the uncharacterized E2 residue requirements. The inability to render the S4520 mutant catalytic in earlier experiments might be explained by its dynamic nature and the absence of a general base that could suppress the acid dissociation constant (pK_a) of the otherwise fully protonated S4520 side chain. Hence, native catalytic activity at C4520 is likely to arise from the intrinsic nucleophilicity of sulfhydryl groups (Extended Data Fig. 8a).

HOIP in complex with E2 (PDB ID: 5EDV; the ubiquitin linked to E2 has been omitted owing to a direct clash with the tandem cysteine domain¹⁸) allows modelling of the E2 into our structure (grey cartoon representation). **e**, Modelling of the proposed ubiquitin relay mechanism. Tandem-cysteine-domain residues and mediator-loop residues from the experimentally determined structure are shown in orange and dark brown, respectively; tandem-cysteine-domain and mediator-loop residues from the model are shown in light orange and mauve, respectively. The modelled E2 is shown as grey cartoon, whereas ubiquitin residues (G75–G76) are shown in blue ball and stick representation.

Although non-lysine ubiquitination has been reported^{25–27}, an E3 ligase that preferentially carries out this function in humans has remained elusive. Our characterization of the novel RCR E3 ligase in MYCBP2 suggests that ubiquitination by esterification is intrinsic to higher eukaryotes, and may be a regulator of synapse development and axon degeneration. Furthermore, non-protein ubiquitin substrates (such as lipids or carbohydrates) have not been reported, but considering the high esterification activity of MYCBP2 towards small-molecule hydroxy compounds, they remain a possibility. It is not immediately clear why the proposed relay (Fig. 6a) mechanism would have evolved. However, transthiolation is a cofactor-independent process that provides an efficient means of shuttling ubiquitin throughout the ubiquitin system¹. We speculate that, on steric grounds, direct E2–E3 transthiolation with the structurally rigid and highly conserved E2 ubiquitin-conjugating domain (Ubc)²⁸, and serine and threonine activity, are mutually exclusive at the esterification site; evolution of the mediator loop addresses this compatibility issue. Bioinformatic analysis revealed that orthologues of MYCBP2 are found in virtually all

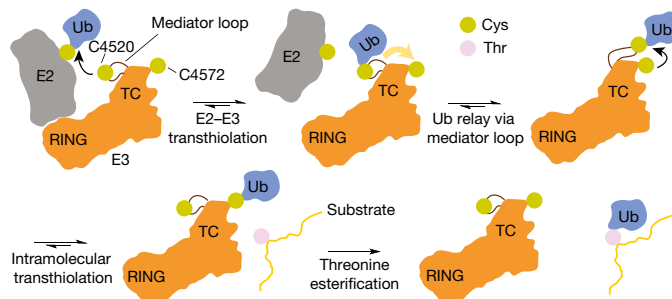


Fig. 6 | Schematic representation of the proposed model of the RCR E3 ligase mechanism. MYCBP2 shows esterification activity towards both serine and threonine, but as it demonstrates a preference for threonine, only this amino acid is shown. TC, tandem cysteine domain.

animals, but human homologues are unlikely to exist (Extended Data Table 2). Stabilization of NMNAT2 through inhibition of MYCBP2 is a promising therapeutic strategy for mitigating neuron damage after injury and administration of chemotherapeutics^{10,29,30}, and for slowing the progression of a range of neurodegenerative diseases including Alzheimer's disease and Parkinson's disease²⁹. The delineation of this apparent ubiquitin relay mechanism and the structural characterization of the molecular machinery responsible suggests new potential targets for treating a range of neurological conditions.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0026-1>.

Received: 4 June 2017; Accepted: 7 March 2018;

Published online: 11 April 2018

- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc. Natl Acad. Sci. USA* **92**, 2563–2567 (1995).
- Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
- Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBC7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
- Hewings, D. S., Flygare, J. A., Bogoy, M. & Wertz, I. E. Activity-based probes for the ubiquitin conjugation–deconjugation machinery: new chemistries, new tools, and new insights. *FEBS J.* **284**, 1555–1576 (2017).
- Pao, K. C. et al. Probes of ubiquitin E3 ligases enable systematic dissection of parkin activation. *Nat. Chem. Biol.* **12**, 324–331 (2016).
- Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).
- Borodovsky, A. et al. Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
- Zhen, M., Huang, X., Bamber, B. & Jin, Y. Regulation of presynaptic terminal organization by *C. elegans* RPM-1, a putative guanine nucleotide exchanger with a RING-H2 finger domain. *Neuron* **26**, 331–343 (2000).
- Wan, H. I. et al. Highwire regulates synaptic growth in *Drosophila*. *Neuron* **26**, 313–329 (2000).
- Grill, B., Murphey, R. K. & Borgen, M. A. The PHR proteins: intracellular signaling hubs in neuronal development and axon degeneration. *Neural Dev.* **11**, 8 (2016).
- Byrne, R., Mund, T. & Licchesi, J. D. F. Activity-based probes for HECT E3 ubiquitin ligases. *ChemBioChem* **18**, 1415–1427 (2017).
- Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E. & Rittinger, K. LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO Rep.* **13**, 840–846 (2012).
- You, J. & Pickart, C. M. A HECT domain E3 enzyme assembles novel polyubiquitin chains. *J. Biol. Chem.* **276**, 19871–19878 (2001).
- Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120 (2012).
- Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **19**, 876–883 (2012).
- Lechtenberg, B. C. et al. Structure of a HOIP/E2–ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature* **529**, 546–550 (2016).
- Dove, K. K. et al. Structural studies of HHARI/UbcH7–Ub reveal unique E2–Ub conformational restriction by RBR RING1. *Structure* **25**, 890–900 (2017).
- Xiong, X. et al. The Highwire ubiquitin ligase promotes axonal degeneration by tuning levels of Nmnat protein. *PLoS Biol.* **10**, e1001440 (2012).
- Babetto, E., Beirowski, B., Russler, E. V., Milbrandt, J. & DiAntonio, A. The Phr1 ubiquitin ligase promotes injury-induced axon self-destruction. *Cell Rep.* **3**, 1422–1429 (2013).
- Milde, S., Gilley, J. & Coleman, M. P. Subcellular localization determines the stability and axon protective capacity of axon survival factor Nmnat2. *PLoS Biol.* **11**, e1001539 (2013).
- Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl–UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
- Olsen, S. K. & Lima, C. D. Structure of a ubiquitin E1–E2 complex: insights to E1–E2 thioester transfer. *Mol. Cell* **49**, 884–896 (2013).
- Cadwell, K. & Coscoy, L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127–130 (2005).
- Shimizu, Y., Okuda-Shimizu, Y. & Hendershot, L. M. Ubiquitylation of an ERAD substrate occurs on multiple types of amino acids. *Mol. Cell* **40**, 917–926 (2010).
- Wang, X., Herr, R. A. & Hansen, T. H. Ubiquitination of substrates by esterification. *Traffic* **13**, 19–24 (2012).
- Alpi, A. F., Chaugule, V. & Walden, H. Mechanism and disease association of E2-conjugating enzymes: lessons from UBE2T and UBE2L3. *Biochem. J.* **473**, 3401–3419 (2016).
- Conforti, L., Gilley, J. & Coleman, M. P. Wallerian degeneration: an emerging axon death pathway linking injury and disease. *Nat. Rev. Neurosci.* **15**, 394–409 (2014).
- Ali, Y. O., Bradley, G. & Lu, H. C. Screening with an NMNAT2-MSD platform identifies small molecules that modulate NMNAT2 levels in cortical neurons. *Sci. Rep.* **7**, 43846 (2017).

Acknowledgements We thank D. Campbell and J. Varghese of the MRC PPU Proteomics Facility, the MRC PPU DNA Sequencing Facility, the European Synchrotron Radiation Facility (ESRF); J. Hastie, H. McLauchlan and F. Brown from MRC PPU Reagents and Services for MYCBP2 antibody production; F. Zuccotto and I. Gilbert for software access; and H. Walden and R. T. Hay for reading of the manuscript and suggestions. This work was funded by the Scottish Funding Council, the UK Medical Research Council (MC_UU_12016/8), BBSRC (BB/P003982/1), and pharmaceutical companies supporting the Division of Signal Transduction Therapy (Boehringer-Ingelheim, GlaxoSmithKline and Merck KGaA). D.M.F.v.A. is funded by a Wellcome Trust Investigator Award (110061).

Author contributions S.V. and K.-C.P. designed research. K.-C.P. carried out experiments with assistance from S.V. N.T.W. performed cloning. A.K. prepared E1, E2 panel and recombinant NMNAT2 protein. K.R. mounted crystals and collected synchrotron diffraction data. M.S. contributed to project conception and carried out molecular synthesis. P.D.M. performed structural modelling, R.S. carried out SEC–MALS. K.H. performed bioinformatics analysis. D.M.F.v.A. solved the MYCBP2(cat) crystal structure. S.V. coded PERL scripts, processed mass spectrometry data and wrote the manuscript with input from all authors.

Competing interests S.V., K.-C.P. and M.S. are authors on patents relating to work presented in this article.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0026-1>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0026-1>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.V.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

General materials. All DNA constructs were verified by DNA sequencing (Medical Research Council Protein Phosphorylation and Ubiquitylation Unit, University of Dundee). DNA for bacterial protein expression was transformed into *Escherichia coli* BL21-DE3 (Merck). All cDNA plasmids and antibodies generated for this study are available on request through our reagents website (<https://mrcppureagents.dundee.ac.uk/>). All solvents and reagents were purchased from Sigma-Aldrich or VWR unless otherwise stated.

Biotin-functionalized ABP preparation. Ubiquitin with a GCSSG N-terminal extension was expressed from plasmid pTXB1-UbΔ74–76-T3C⁷. An equivalent plasmid encoding ubiquitin residues 1–74 (pTXB1-UbΔ75–76-T3C) was also created. Ubiquitin thioesters were obtained as described previously, generating cysteine tagged Cys-Ub_{1–73}-SR and Cys-Ub_{1–74}-SR, respectively⁷, where SR is SCH₂CH₂SO₃H. The extended Ub_{1–74} was included as it retains Arg74, which forms a favourable electrostatic interaction with the RBR E3 HOIP³¹. Cys-Ub_{1–73}-SR (30 mg) was reconstituted by the addition of DMSO (116 μl) followed by H₂O (456 μl). An aqueous stock solution (48 mM) of EZ-link Iodoacetyl-PEG2-biotin (ThermoFisher) was prepared and 200 μl was added to the Cys-Ub_{1–73}-SR solution (580 μl) followed by the addition of 900 μl degassed buffer (50 mM Na₂HPO₄ pH 7.5, 150 mM NaCl). The reaction was incubated at 23 °C for 1 h and monitored by LC-MS. The protein (biotin-Ub_{1–73}-SR) was then further purified by semi-preparative reversed-phase high-pressure liquid chromatography (RP-HPLC) (Column: BioBasic-4; part number: 72305-259270). A gradient of 20% buffer A to 50% buffer B was applied at a flow rate of 10 ml min⁻¹ over 60 min (buffer A, 0.1% TFA in H₂O; buffer B, 0.1% TFA in acetonitrile). The above procedure was repeated to generate Biotin-Ub_{1–74}-SR. HPLC fractions containing biotin-Ub_{1–73}-SR were pooled and lyophilized (yield: biotin-Ub_{1–73}-SR, 75–85%; biotin-Ub_{1–74}-SR, 40–50%) (Extended Data Fig. 1a, d). Biotin-tagged ABPs containing thioacrylamide warheads were then prepared as previously described⁷, using the E2-recognition elements UBE2D2*, UBE2D2*(F62A), UBE2L3* and UBE2L3*(F63A), resulting in ABP1, ABP3, ABP2 and ABP4, respectively (Extended Data Fig. 1b, c, e, f); the asterisk denotes E2 in which non-catalytic Cys residues were mutated to Ser. ABPs based on UBE2D2* and UBE2D3* bearing hexahistidine reporter tags and thioacrylamide warheads (Extended Data Fig. 3a) were also prepared, yielding ABP5 and ABP6, respectively.

Cell culture and lysis. SH-SY5Y cells were cultured and lysed as previously described⁷. HEK293 cells were cultured (37 °C, 5% CO₂) in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 2.0 mM L-glutamine and antibiotics (100 units ml⁻¹ penicillin, 0.1 mg ml⁻¹ streptomycin). Cell transfections were performed using polyethylenimine (Polysciences) according to the manufacturer's instructions. MG-132 (50 μM) was added to cells two hours before lysis. Cells were rinsed with ice-cold PBS and extracted in lysis buffer (1% NP-40, 50 mM Tris-HCl pH 7.5, 1 mM EGTA, 1 mM EDTA, 0.27 M sucrose, 10 mM sodium 2-glycerophosphate, 0.2 mM phenylmethanesulfonyl fluoride (PMSF), 1.0 mM benzamide, 1 mM sodium orthovanadate, 50 mM sodium fluoride and 5 mM sodium pyrophosphate, 50 mM iodoacetamide and cComplete EDTA-free protease inhibitor cocktail (Roche)). Lysates were then clarified by centrifugation at 4 °C for 30 min at 21,100 g. Supernatants (total cell extracts) were collected and protein concentration was determined by Bradford assay. For the base-lability test, indicated cell lysates were further incubated with 0.5 M hydroxylamine, pH 9.0 at 37 °C for 30 min. Mycoplasma tests were carried out in accordance with departmental protocols and the results were negative.

Immunoblotting. Samples were mixed with NuPAGE LDS sample buffer (ThermoFisher) without boiling, and resolved by SDS-PAGE (4–12% NuPage gel, ThermoFisher) with MOPS or MES running buffer and transferred on to 0.45-μm nitrocellulose membranes (GE Life Sciences). Membranes were blocked with PBS-T buffer (PBS, 0.1% Tween-20) containing 5% (w/v) non-fat dried skimmed milk powder (PBS-TM) at room temperature for 1 h. Membranes were subsequently probed with the indicated antibodies in PBS-T containing 5% (w/v) bovine serum albumin (BSA) overnight at 4 °C. Detection was performed using horseradish peroxidase (HRP)-conjugated secondary antibodies in PBS-TM for 1 h at 23 °C. ECL western blotting detection reagent (GE Life Sciences) was used for visualization according to the manufacturer's protocol.

Antibodies. His-tagged species were probed with 1:10,000 anti-His primary antibody (Clontech, #631212). Alpha tubulin (1E4C11), mouse mAb (Proteintech) was used at 1:10,000 dilution. The MYCBP2 antibody was raised in sheep inoculated with residues 4,378–4,640 of human MYCBP2 by MRC PPU Reagents and Services. Anti-MYCBP2 from the second bleed of SA357 was affinity-purified against the antigen and used at 0.5 μg ml⁻¹. Mouse monoclonal NMNAT2 antibody (clone 2E4, Sigma Aldrich) was used at 0.5 μg ml⁻¹.

Activity-based proteomic profiling of SH-SY5Y cells. SH-SY5Y total cell lysate (4.5 mg, 550 μl) was mixed with ABPs 1, 2, 3 or 4 (3 μM) and incubated at 30 °C for 4 h. To induce Parkin activation, cells were administered with oligomycin (5 μM) and antimycin A (10 μM) for 3 h. Control enrichments were also performed

without probes. Extracts were mixed with 100 μl Pierce Streptavidin Plus UltraLink Resin (ThermoFisher Scientific) and diluted with 6% SDS solution (20 μl) to a final concentration of 0.2%. Samples were incubated for 4 h at 4 °C and resin washed (2 ml 0.2% SDS in PBS; 2 ml PBS; 1 ml 4 M urea in PBS; 2 ml PBS) and then resuspended in 190 μl Tris buffer (50 mM Tris pH 8, 1.5 M urea). Resin-bound proteins were reduced with TCEP (5 mM) for 30 min at 37 °C and then alkylated with iodoacetamide (10 mM) at 23 °C for 20 min. DTT (10 mM) was then added followed by washing with buffer (50 mM Tris pH 8, 1.5 M urea) to a final volume of 300 μl. Trypsin (2 μg) was then added and further incubated at 37 °C for 14 h. Trifluoroacetic acid was added to a final concentration of 0.1% and samples were desalted with a C18 MacroSpin column (The Nest Group). LC-MS/MS analysis was performed on an LTQ Orbitrap Velos instrument (Thermo Scientific) coupled to an Ultimate Nanoflow HPLC system (Dionex). A gradient running from 3% solvent B to 99% solvent B over 345 min was applied (solvent A, 0.1% formic acid and 3% DMSO in H₂O; solvent B, 0.08% formic acid and 3% DMSO in 80% MeCN).

Data processing. Raw files were searched against the Swiss-Prot database and a decoy database using the MASCOT server (Matrix Science). Trypsin specificity with up to three missed cleavages was applied. Cysteine carbamylation was set as a fixed modification and variable modifications were methionine oxidation and dioxidation. A PERL script was used to extract the number of rank 1 peptides for each protein from the MASCOT search results and this figure was used as the number of spectral counts. A second PERL script filtered the data by searching the human swisspfam_v30 database using the E3 domain terms RING, HECT, IBR and zf-UBR. Manual curation was also carried out which involved the addition of E1 enzymes. Any proteins with fewer than three spectral counts and less than 14-fold spectral-count enrichment relative to control experiments in which probes were omitted from the list. Pairwise datasets were then plotted as column charts in Prism (GraphPad Software).

Cloning of MYCBP2(cat). Human MYCBP2 (NM_015057.4) sequences were amplified from full-length Addgene plasmid no. 2570. Wild-type and mutant fragments were subcloned as BamHI–NotI inserts into pGEX6P-1 (GE Life Sciences) for bacterial expression, or a modified version of pcDNA TM5/FRT/TO (ThermoFisher) containing an N-terminal Myc tag for mammalian expression.

UBE1 and E2 expression and purification. His₆-UBE1 was expressed in Sf21 cells and purified via its tag as previously described³². Phosphate-buffered saline was used throughout the purification and hydroxy-containing compounds were avoided. UBE2D3 was expressed as an N-terminally His₆-tagged protein in BL21 cells, purified over Ni-NTA-agarose and dialysed into 50 mM Na₂HPO₄ pH 7.5, 150 mM NaCl, 0.5 mM TCEP. UBE2A was expressed as a GST fusion in *E. coli* and the GST tag was proteolytically removed. The other E2s were expressed as recombinant bacterial proteins and purified via their His tags and buffer exchanged by size exclusion chromatography into running buffer (50 mM Na₂HPO₄ pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 0.015% Brij-35) using a Superdex 75 column (GE Life Sciences).

Expression and purification of MYCBP2 and GST-MYCBP2. Wild-type and mutant GST-tagged MYCBP2(cat) were expressed at 16 °C overnight and purified with glutathione resin (Expadeon) using standard procedures. GST-tagged constructs were eluted with glutathione and untagged constructs were obtained by on-resin cleavage with rhinovirus 3C protease. Proteins were buffer exchanged into storage buffer (50 mM Na₂HPO₄ pH 7.5, 150 mM NaCl, 1.0 mM TCEP) and kept at –80 °C.

Expression and purification of NMNAT2. NMNAT2 was expressed with a His₆-SUMO tag in BL21(DE3) cells, induced with 0.1 mM IPTG and incubated for expression at 16 °C. The cells were collected and lysed in 50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 0.2 mM EGTA, 20 mM imidazole, 20 mM L-arginine, 0.015% Brij-35, 1 mM leupeptin, 1 mM Pefabloc, 1 mM DTT using standard protocols and the protein was purified over Ni-NTA-agarose. The eluted protein was incubated with His-SENPI protease during dialysis against PBS, 20 mM L-arginine, 1 mM DTT. The tag and protease were depleted against Ni-NTA-agarose and NMNAT2 was concentrated and subjected to chromatography on a Superdex 75h 10/30 column into PBS, 20 mM L-arginine.

Activity-based protein profiling of MYCBP2 cysteine mutants. The indicated MYCBP2 mutant was diluted into Tris buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl) to a final concentration of 3 μM. ABP6 was added (12 μM) and incubated with E3 ligase at 30 °C for 4 h. Reactions were quenched by the addition of 4 × LDS loading buffer (supplemented with ~680 mM 2-mercaptoethanol) and samples were resolved by SDS-PAGE (4–12% NuPage gel) followed by Coomassie staining or anti-His immunoblotting.

Trypic MS/MS sequencing of probe-labelled MYCBP2. Crosslinking mass spectrometry using ABP6 was carried out as previously described⁷. In summary, the Coomassie-stained SDS-PAGE band corresponding to ABP-labelled wild-type MYCBP2 was analysed by LC-MS/MS using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) coupled to an Ultimate Nanoflow HPLC system (Dionex). A gradient running from 0% solvent A to 60% solvent B over

120 min was applied (solvent A = 0.1% formic acid in H₂O; solvent B = 0.08% formic acid in 80% MeCN). Fragment ions were generated by HCD and 1⁺, 2⁺ and 3⁺ precursor ions were excluded. Raw data were searched using the pLink software³³ against UBE2D3* and MYCBP2 sequences with trypsin specificity (up to two missed cleavages). The error window for MS/MS fragment ion mass values was set to the software default of 20 p.p.m. A crosslinker monoisotopic mass of 306.1805 Da was manually added, which accounted for the theoretical mass difference associated with formation of a bis(thioether) between two Cys residues derived from ABP6, which was based on UBE2D3* and contained a thioacrylamide AVS warhead⁷.

Tris-glycerol-mediated E2-discharge assay. Assays were carried out in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5 mM MgCl₂ containing the indicated MYCBP2 mutant (15 μ M), UBE1 (1.5 μ M), UBE2D3 (15 μ M), ubiquitin (37 μ M) and ATP (10 mM). The reactions were incubated at 37 °C for 30 min. Reactions were terminated by the addition of 4 \times LDS loading buffer (with and without ~680 mM 2-mercaptoethanol). A C4572S sample was further incubated with 0.14 N NaOH at 37 °C for 20 min and samples were resolved by SDS-PAGE (4–12% NuPage gel) and visualized by Coomassie staining.

LC-MS analysis of nucleophile discharge assays. Reactions were prepared as described for the E2 discharge assay. After 30 min, the reaction was analysed using an Agilent 1200/6130 LC-MS system (Agilent Technologies) using a 10–75% gradient over 20 min (buffer A, H₂O + 0.05% TFA; buffer B, acetonitrile + 0.04% TFA).

Preparation of Cy3B-Ub. Ubiquitin bearing a TEV protease-cleavable N-terminal hexahistidine tag followed by an ACG motif was expressed in bacteria from a pET plasmid (R. Hay, University of Dundee). Protein was purified by Ni-affinity chromatography, cleaved from the tag with TEV protease then buffer exchanged into reaction buffer (50 mM HEPES, pH 7.5, 0.5 mM TCEP). Protein was concentrated to 2 mg ml⁻¹ and 221 μ l (50 nmol) of this was mixed with Cy3B-maleimide (150 nmol, GE Life Sciences) in a final volume of 300 μ l and agitated for 2 h at 25 °C. Labelled protein was then further purified with a P2 Centri-Pure desalting column (EMP Biotech) with degassed buffer (50 mM Na₂HPO₄, 150 mM NaCl).

MYCBP2 thioester-ester trapping assay. UBE1 (2 μ M) was mixed with Cy3B-Ub (1 μ M) in 40 mM Na₂HPO₄-HCl pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5 mM MgCl₂ (Fig. 2e, lanes 1, 2). The reaction was then initiated by the addition of ATP (5 mM) and incubated for 10 min at 25 °C. Samples (lanes 3, 4) were taken and combined with UBE2D3 (10 μ M). After a further 10 min at 25 °C, samples (lanes 5, 6) were taken and combined with GST-MYCBP2(cat) (wild-type, C4520S, C4520A, C4572S, C4572A, C4520A/C4572S or C4520S/C4572A) (15 μ M). The reactions were incubated at 25 °C for 30 s and terminated by the addition of 4 \times LDS loading buffer (either non-reducing or reducing). For Ub-GST-MYCBP2(cat-C4572S) ester bond cleavage, 0.14 N NaOH was added after E3 reaction with E1 and E2 for 30 s and then further incubated at 37 °C for 20 min. The gel was then scanned with a Chemidoc Gel Imaging System (BioRad).

Multiple-turnover amino acid and peptide panel discharge assays. Stock solutions (0.5 M) of amino acids were dissolved in Milli-Q water and pH was adjusted to ~8. Peptides with the sequence Ac-EGXGN-NH₂ (where X = K, S or T) were obtained from Bio-Synthesis. Stock peptide solutions (200 mM) were dissolved in Milli-Q water and pH was adjusted to ~8. An E2 (UBE2D3)-charging reaction was carried out in 40 mM Na₂HPO₄-HCl pH 8.0, 150 mM NaCl, 0.5 mM TCEP containing UBE1 (250–500 nM), UBE2D3 (20 μ M), ubiquitin (50 μ M) or Cy3B-Ub (25 μ M), MgCl₂ (5 mM) and ATP (10 mM). The reaction was incubated at 37 °C for 15 min and then equilibrated to 23 °C for 3 min. An equivalent volume of nucleophile sample containing small molecule-peptide nucleophile (100 mM) and GST-MYCBP2 (10 μ M) was then added and incubated at 23 °C. Samples were taken at the specified time points and analysed as described for Tris-glycerol-mediated E2 discharge assay.

Cy3B-Ub was visualized using a Chemidoc Gel Imaging System (Bio-Rad). LC-MS was carried out as described for Tris-glycerol discharge but amino acid substrate samples were quenched by the addition of 2:1 parts quenching solution (75% acetonitrile, 2% TFA) and peptide substrate samples were quenched by addition of 1:1 parts quenching solution.

Multiple-turnover E2 discharge panel. E2s were screened for threonine discharge activity with GST-MYCBP2_{cat} as described for the amino acid panel. E2s were also incubated in the presence of threonine but in the absence of GST-MYCBP2_{cat}. These samples provided a reference to distinguish between intrinsic E2-Ub instability and E3-dependent discharge.

Single-turnover E2 mutant discharge by in-gel fluorescence. E2 mutants^{16,17,34–36} (10 μ M) were charged with Cy3B-labelled ubiquitin (12.5 μ M) in a final volume of 12 μ l at 37 °C for 20 min then cooled at 23 °C for 3 min. E2 recharging was then blocked by the addition of MLN4924 derivative, compound 1 (25 μ M)³⁷, which inhibits E1, and then incubated for a further 15 min. The mixture was then mixed with 12 μ l of GST-MYCBP2_{cat} (5 μ M) and threonine (100 mM) and incubated at 23 °C for the specified time. Analysis was carried out as for multiple-turnover assays. To account for intrinsic E2-Ub instability the mean percentage discharge

($n = 2$) was calculated against a parallel incubation without E3. Data were plotted using Prism (GraphPad).

Expression and purification of ARIH1 and UBE3C. ARIH1 (residues 1–394) (Dundee clone DU24260) was expressed as an N-terminally GST-tagged fusion protein in BL21 cells. UBE3C (residues 641–1,083) (Dundee Clone DU45301) was expressed as an N-terminally GST-tagged fusion protein in Sf21 cells using a baculovirus infection system.

Calculation of observed rate constants for E3-substrate-dependent single-turnover E2-Ub discharge. UBE2D3 or UBE2L3 (5 μ M) were charged with Cy3B-labelled ubiquitin (8 μ M) in a final volume of 30 μ l at 37 °C for 25 min and then incubated at 23 °C for 3 min. Single-turnover conditions for E2-Ub discharge were achieved by E1 inhibition with MLN4924 derivative, compound 1 (25 μ M) and then incubated for a further 15 min. The mixture was then mixed with 30 μ l of MYCBP2(cat) or ARIH1(1–394) (HHARI) or UBE3C(641–1083) (1 μ M) and threonine (100 mM) and incubated at 23 °C for the specified time. Samples were quenched with non-reducing 4 \times LDS loading buffer and resolved by SDS-PAGE (Bis-Tris 4–12%). The gel was then scanned with a Chemidoc Gel Imaging System (Bio-Rad) and subsequently Coomassie stained. E2-Ub signals were quantified using Fiji software. Observed rate constants were obtained by fitting reaction progress curves to a single exponential function using Prism (GraphPad).

MYCBP2 crystallization. MYCBP2 was expressed as described for untagged protein. After protease cleavage of the tag the protein was further purified by size-exclusion chromatography using an ÄKTA FPLC system and a HiLoad 26/600 Superdex 75 pg column (GE Life Sciences). The running buffer consisted of 20 mM HEPES pH 7.4, 150 mM NaCl, 4 mM DTT. Combined fractions were concentrated to 10.4 mg ml⁻¹. Sparse matrix screening was carried out and bipyrimalid crystals were obtained from the Morpheus screen condition C1 (Molecular Dimensions). A subsequent optimization screen yielded multiple crystals (Buffer system 1 (MES/imidazole) pH 6.7, 23.3 mM Na₂HPO₄, 23.3 mM (NH₄)₂SO₄, 23.3 mM NaNO₃, 18% PEG500 MME, 9% PEG20000). A single crystal was soaked in mother liquor and further cryoprotected by supplementation with 5% PEG400 and frozen in liquid N₂. Data were collected to 1.75 Å at the European Synchrotron Radiation Facility at Beamline ID23-1. Energy was set to the peak value of 9.669 keV (1.2823 Å), as determined by an absorption edge energy scan. A total of 360° were collected with an oscillation range of $\Omega = 0.1^\circ$. The phase problem was solved by locating six Zn²⁺ sites in the anomalous signal and solvent flattening with the SHELX suite. An initial model was built by ARP/wARP³⁸ and subsequently optimized by manual building in COOT³⁹ and refinement with REFMAC5⁴⁰, resulting in the final model with statistics as shown in Extended Data Fig. 7. Final Ramachandran statistics were favoured: 95.55%, allowed: 3.24%, outliers: 1.21%.

Size exclusion chromatography with multi-angle light scattering (SEC-MALS). SEC-MALS experiments were performed on an Ultimate 3000 HPLC system (Dionex) with an in-line miniDAWN TREOS MALS detector and Optilab T-rEX refractive index detector (Wyatt). In addition, the elution profile of the protein was also monitored by UV absorbance at 280 nm. A Superdex 75 10/300 GL column (GE Life Sciences) was used. Buffer conditions were 50 mM Na₂HPO₄ pH 7.5, NaCl 150 mM, 1.0 mM TCEP and a flow rate of 0.3 ml min⁻¹ was applied. Sample (50 μ l, 5.5 mg ml⁻¹) was loaded onto the column with a Dionex autosampler. Molar masses spanning elution peaks were calculated using ASTRA software v.6.0.0.108 (Wyatt).

Mediator loop modelling. Mediator loop residues were built and geometry optimized within the Bioluminate Software (Schrödinger). Side chains were modified within COOT³⁹ and figures were generated with PyMOL (Schrödinger). Ramachandran analysis was carried with the RAMPAGE server⁴¹.

NMNAT2 ubiquitination assay. NMNAT2 (5 μ M) was mixed with E1 (500 nM), UBE2D3 (10 μ M), MYCBP2(cat) (10 μ M), ubiquitin (50 μ M), ATP (10 mM) and made up with 10 \times pH 7.5 buffer (40 mM Na₂H₂PO₄ pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 0.5 mM TCEP). The reactions were incubated at 37 °C for 1 h and terminated by the addition of 4 \times LDS loading buffer (either non-reducing or reducing). For base-lability test, reactions were supplemented with 0.14 N NaOH and then further incubated at 37 °C for 20 min.

Bioinformatic analysis. Proteins belonging to the RCR family were identified by generalized profile searches. Overall 671 such sequences were identified. The sequences were aligned by profile-guided alignment using the pftools package. For identifying representative sequences from various taxa, the Belvu program (Sanger Institute) was used to remove sequences with > 80% identity to other sequences. Truncated and misassembled proteins were removed manually, resulting in 130 representative tandem-cysteine-domain sequences.

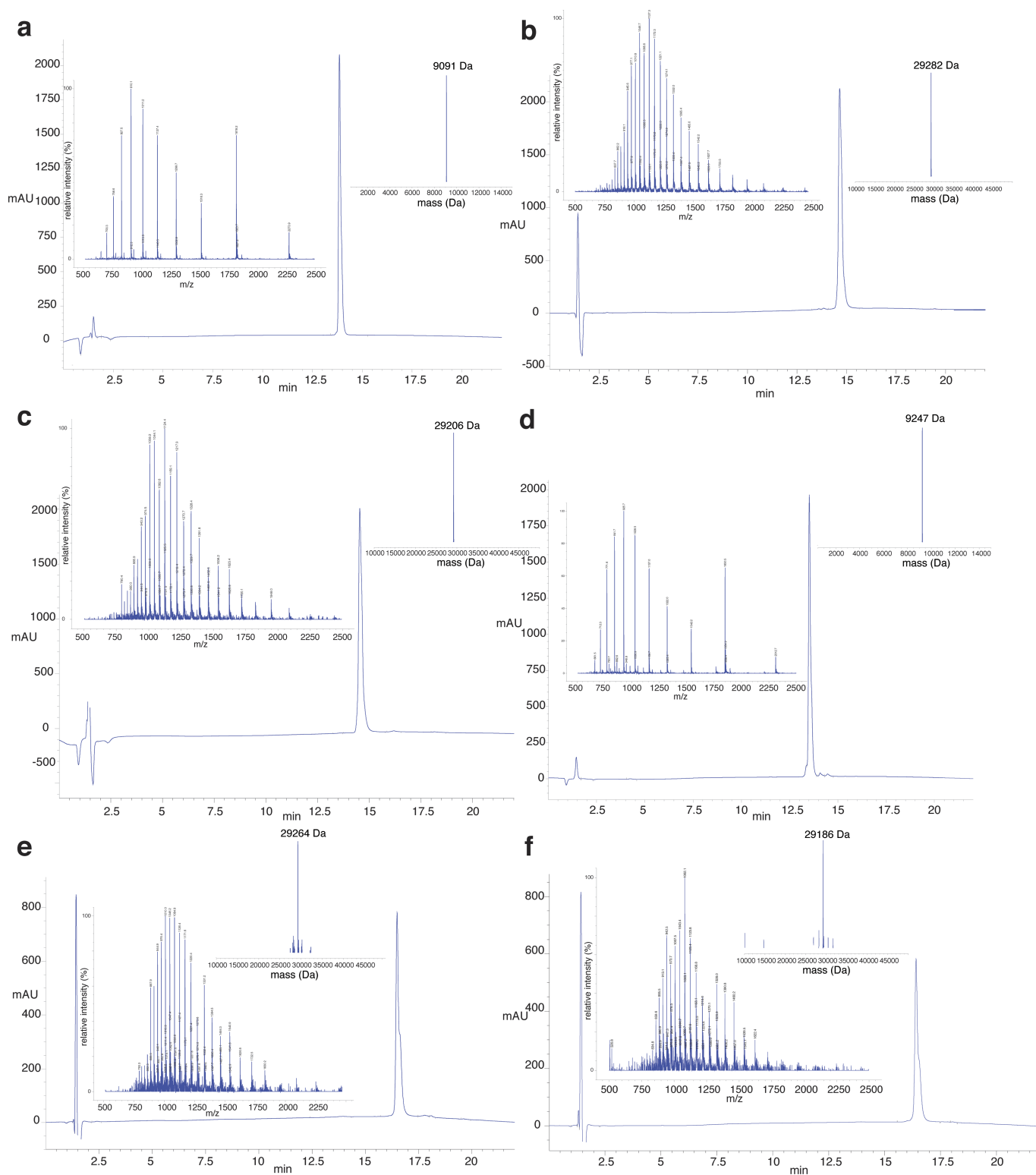
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Custom PERL scripts for processing of mass spectrometry data are available upon request.

Data availability. All data supporting the findings reported here are in the main figures, Extended Data or Supplementary files and information. Coordinates of

the protein structure have been deposited in Protein Data Bank with the accession number 5O6C. For gel source data, see Supplementary Fig. 1.

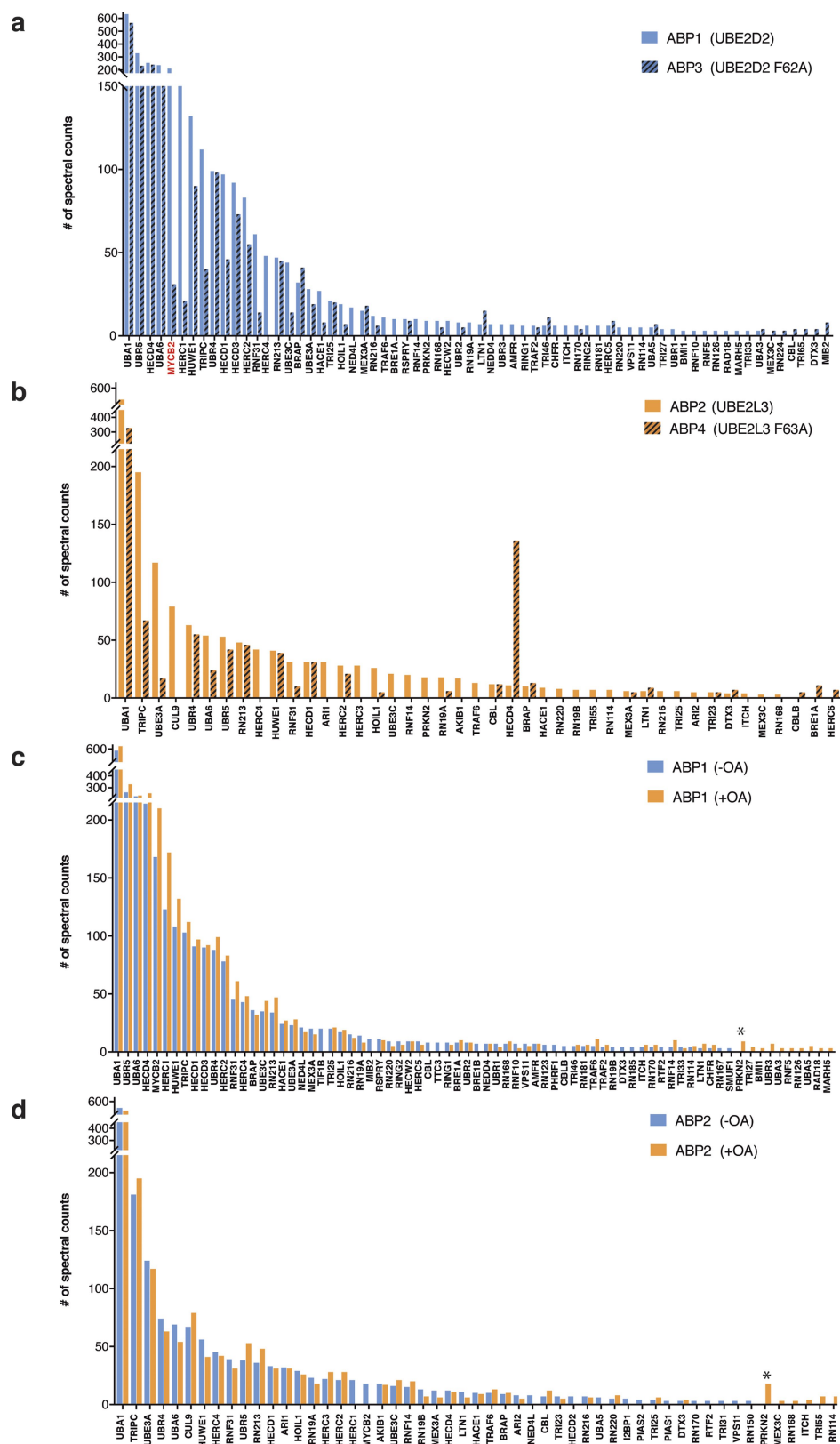
31. Stieglitz, B. et al. Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **503**, 422–426 (2013).
32. Stanley, M. et al. Orthogonal thiol functionalization at a single atomic center for profiling transthiolation activity of E1 activating enzymes. *ACS Chem. Biol.* **10**, 1542–1554 (2015).
33. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
34. Pruneda, J. N. et al. Structure of an E3:E2–Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942 (2012).
35. Wu, P. Y. et al. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–5250 (2003).
36. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat. Struct. Mol. Biol.* **13**, 491–499 (2006).
37. Brownell, J. E. et al. Substrate-assisted inhibition of ubiquitin-like protein-activating enzymes: the NEDD8 E1 inhibitor MLN4924 forms a NEDD8–AMP mimetic in situ. *Mol. Cell* **37**, 102–111 (2010).
38. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
40. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
41. Lovell, S. C. et al. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* **50**, 437–450 (2003).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | LC-MS characterization of biotinylated-ABP intermediates and biotinylated ABPs. a–f, E3s can have distinct E2 preferences⁵, so to obtain broad coverage we prepared biotinylated ABPs based on the promiscuous E2 UBE2D2 (ABP1), and the HECT/RBR-specific E2 UBE2L3 (ABP2). As controls we prepared ABPs containing a point mutation in the E2-recognition component to disrupt or impair E3 ligase binding, UBE2D2(F62A) (ABP3) and UBE2L3(F63A) (ABP4)⁷. Ubiquitin in ABP1 and ABP3 was extended by a single residue relative to that previously reported⁷ (Ub_{1–74} rather than Ub_{1–73}), as this improved labelling efficiency of the RBR E3 HOIP. **a,** Characterization of biotin-labelled, truncated ubiquitin-thioester intermediate, biotin-Ub_{1–73}-SR used for ABP2 and ABP4. HPLC chromatogram monitoring UV absorbance at 214 nm (as for all subsequent intermediates and ABPs). ESI mass spectrum (inset left) and deconvoluted mass spectrum (inset right). Expected mass = 9,093 Da (–Met); found mass = 9,091 Da. **b,** Characterization of UBE2L3 ABP2. ESI mass spectrum (inset left) and

deconvoluted mass spectrum (inset right). Expected mass = 29,286.9 Da (–Met); found mass = 29,282 Da. **c,** Characterization of UBE2L3(F63A) ABP4. ESI mass spectrum (inset left) and deconvoluted mass spectrum (inset right). Expected mass = 29,210.8 Da (–Met); found mass = 29,206 Da. **d,** Characterization of probe intermediate with extended ubiquitin C terminus, biotin-Ub_{1–74}-SR used to make ABP1 and ABP3. ESI mass spectrum (inset left) and deconvoluted mass spectrum (inset right). Expected mass = 9,249.2 Da (–Met); found mass = 9,247 Da. **e,** Characterization of UBE2D2 ABP1. ESI mass spectrum (inset left) and deconvoluted mass spectrum (inset right). Expected mass = 29,268.8 Da (–Met); found mass = 29,264 Da. **f,** Characterization of UBE2D2(F62A) ABP2. ESI mass spectrum (inset left) and deconvoluted mass spectrum (inset right). Expected mass = 29,192.7 Da (–Met); found mass = 29,186 Da. Intermediates and probes have been prepared and characterized more than three times with similar results.



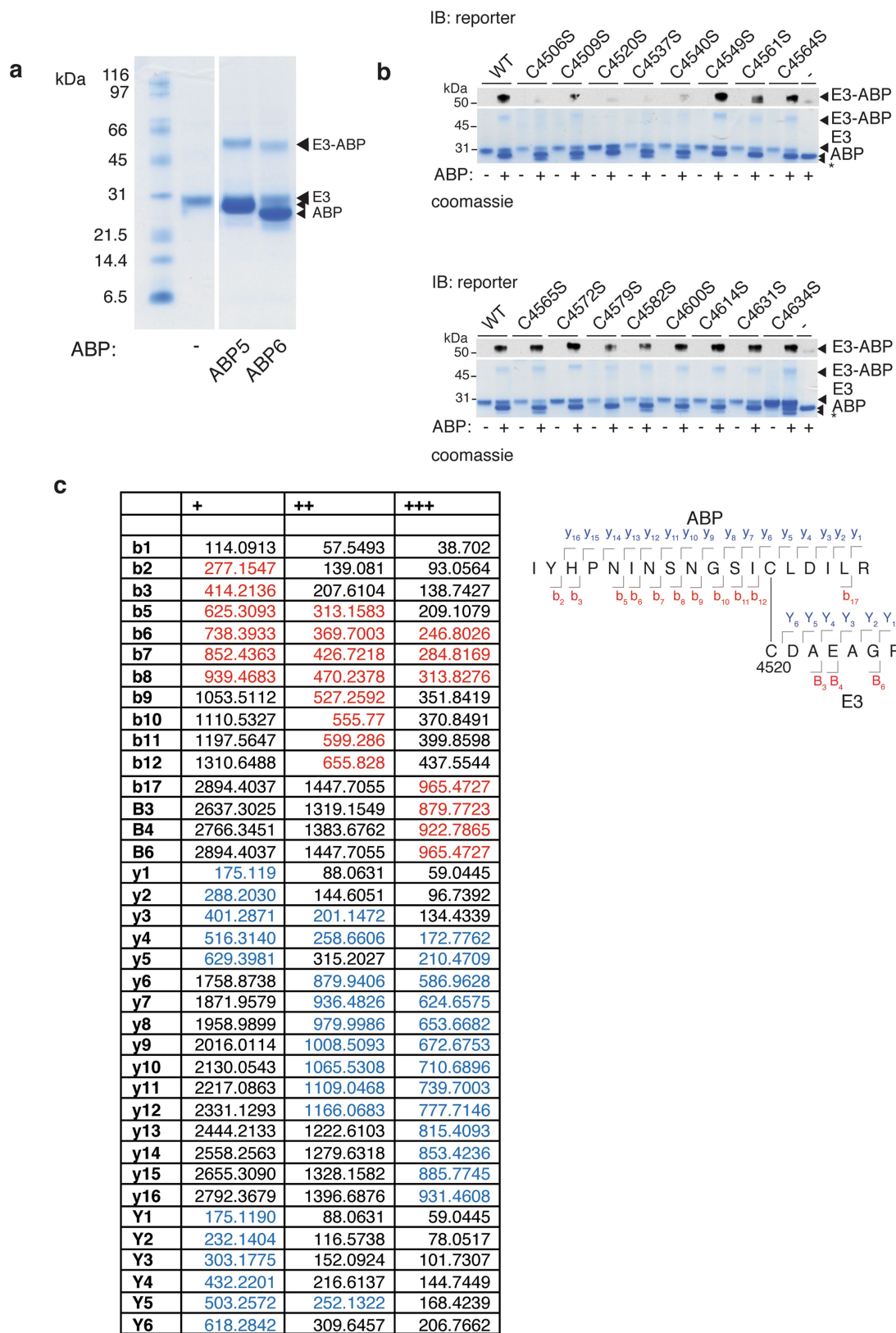
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Activity-based proteomic profiling of

neuroblastoma SH-SY5Y cells. a–d, Parallel profiling of neuroblastoma SH-SY5Y cell extracts was carried out with ABPs 1–4. As an additional control, cells were left untreated or treated with inhibitors of oxidative phosphorylation, oligomycin and antimycin A, which enables activity-dependent labelling of the RBR E3 Parkin⁷. ABP-labelled proteins were enriched against streptavidin resin followed by on-resin tryptic digestion. Obtained peptides were analysed by data-dependent LC–MS/MS.

Recovered proteins were filtered against E3-associated PFAM domain terms (RING, HECT, IBR, zf-UBR) and proteins with fewer than three spectral counts were excluded. E3s that did not demonstrate more than 14-fold spectral count enrichment compared to control purifications, in which ABP was withheld, were also excluded. E1s yielded a strong signal, because they undergo transthiolation and are highly enriched by our ABPs. The number of spectral counts for the majority of HECT/RBR E3s was reduced by > 50% relative to their parental counterpart when the binding-defective control probes ABP3 and ABP4 were used (Fig. 1c). The aggregate number of recovered HECT/RBR E3s from ABP1 and ABP2 was 33 (22 HECT and 11 RBR), representing around 80% of the currently annotated HECT/RBR E3s (Fig. 1d). A subset of E3s remain permissive to control probes ABP3 and ABP4; we cannot establish whether this is because the respective E3s are labelled in an activity-independent manner or whether they are permissive to the F62A or F63A mutation.

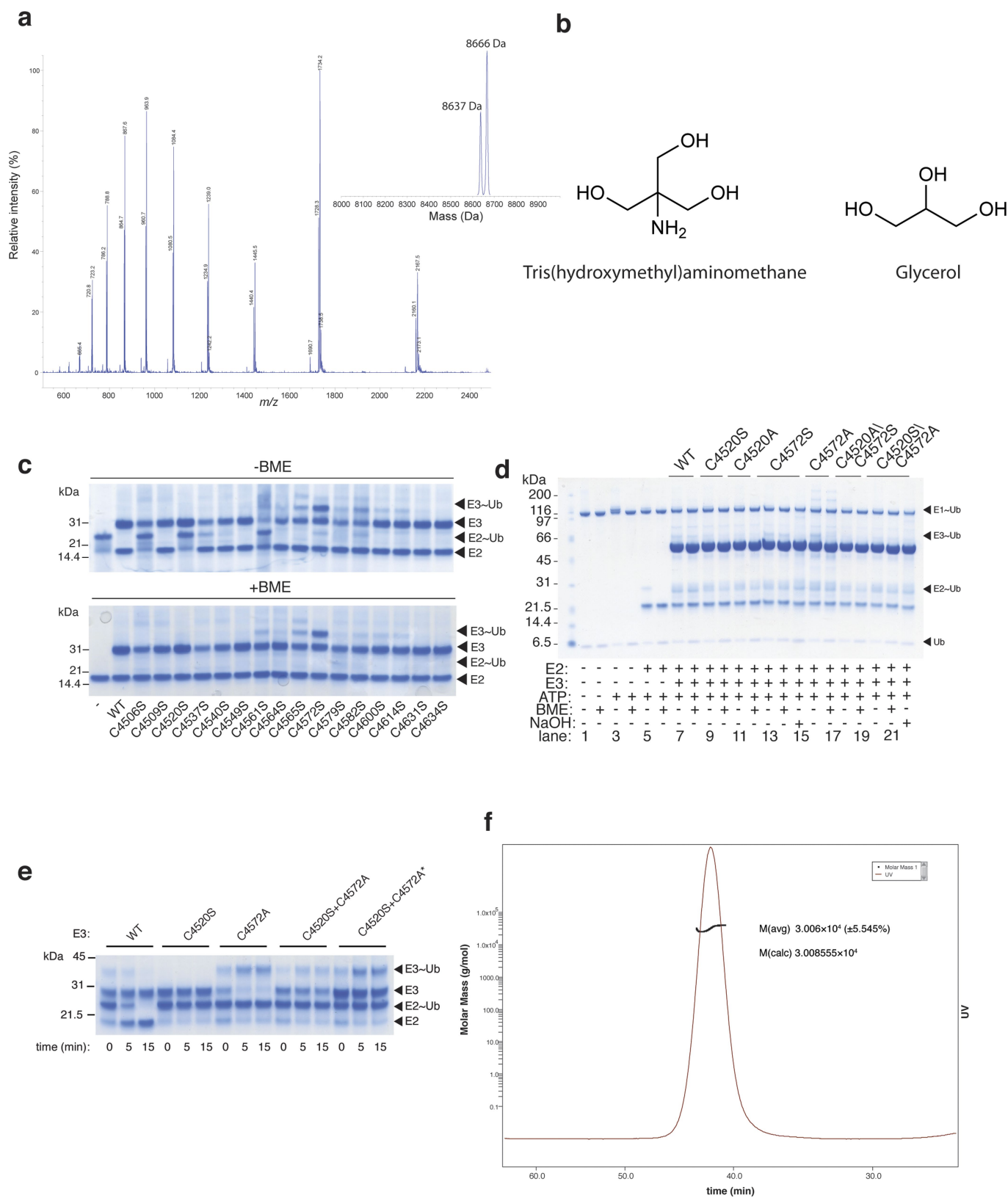
Furthermore, ABP-dependent spectral count signals are not normalized against protein abundance. Therefore, we cannot deconvolute the effects of E3 activation stoichiometry from E3 abundance (that is, highly abundant E3s that are in a low-activation state could yield disproportionately high signals in our data). **a,** The number of spectral counts for the recovered E1 and E3 proteins plotted against protein ID for UBE2D2 ABP1 and its respective control probe UBE2D2(F62A) ABP3. MYCBP2 yields a high signal with ABP1, which is reduced by more than 50% with ABP3. A number of RING E3s that bind the ABP are also labelled, and for the majority of cases this is presumed to be mechanistically off-target labelling that is exacerbated by the high sensitivity of mass spectrometry-based detection. Another possibility is that hitherto undiscovered RING-linked E3s are being labelled. **b,** The number of spectral counts for the recovered E1 and E3 proteins plotted against protein ID for UBE2L3 ABP2 versus the respective control probe UBE2L3(F63A) ABP4. MYCBP2 is not detected with UBE2L3 ABP2 and ABP4. **c,** The number of spectral counts for E1 and E3 proteins obtained with ABP1 for untreated versus oligomycin and antimycin A-treated cells. **d,** The number of spectral counts for E1 and E3 proteins obtained with ABP2 for untreated versus oligomycin and antimycin A-treated cells. Parkin peptides were only recovered from cells treated with oligomycin and antimycin A, consistent with activity-dependent Parkin labelling. Thus, for at least a subset of detected E3s, spectral counts correlate with E3 activity.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | ABPs label MYCBP2(cat) C4520 with high selectivity. **a**, Recombinant MYCBP2(cat) was profiled with His-tagged ABPs based on the E2s UBE2D2 (ABP5) and UBE2D3 (ABP6). The experiment was repeated twice with similar results. **b**, Putative active-site cysteines in MYCBP2 were determined by ABP profiling of a panel of cysteine-to-serine mutants. MYCBP2(cat) mutant (3 μ M) was incubated with ABP6 (12 μ M) at 30 °C for 4 h. ABP-treated samples were resolved by SDS–PAGE and visualized by Coomassie staining and immunoblotting against the hexahistidine reporter tag on the ABP. Mutation of three cysteine residues (C4506, C4520 and C4537) abolished ABP labelling. The asterisk corresponds to inadvertent cleavage of the hexahistidine tag from the ABP due to trace protease contamination of the E3 preparations. The

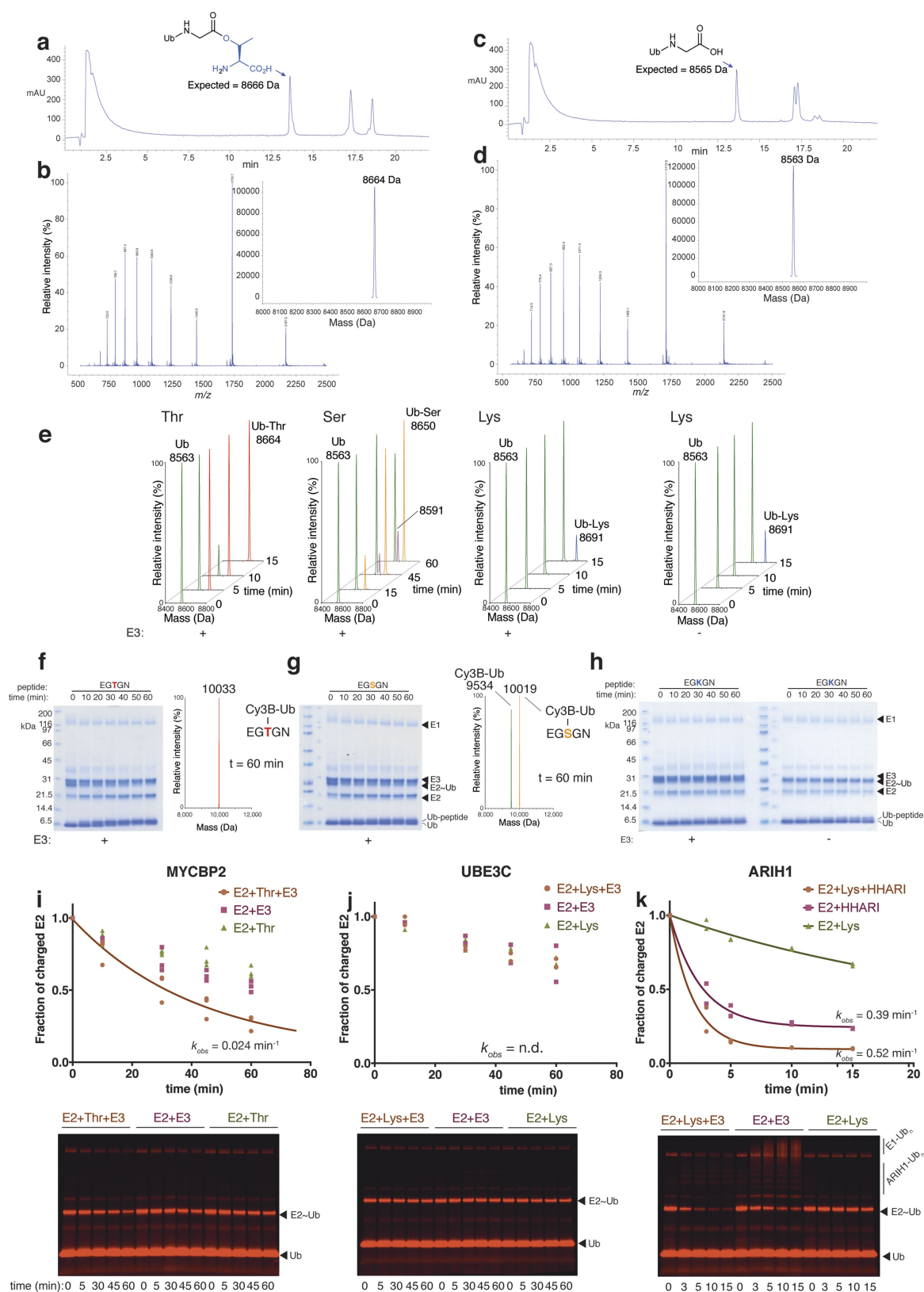
experiment was repeated twice with similar results. **c**, Using the pLink software, 38 spectral matches corresponding to cysteine-labelling sites in wild-type MYCBP2_{cat} were identified. Thirty-six of these corresponded to C4520. One of the two remaining matches corresponded to C4440, a predicted Zn-coordinating residue in the RING domain. The other remaining match corresponded to C4600, which did not significantly affect ABP labelling when mutated. The table lists the predicted and found fragment ions for the representative spectrum depicted in Fig. 2b. The spectrum is for a 5⁺ precursor ion (expected m/z = 614.5094; observed m/z = 614.5088). A mass tolerance of 20 p.p.m. was applied for fragment ion assignment. Experiment was carried out once.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Esterification activity of MYCBP2(cat) and further data in support of a dual cysteine mechanism that operates in cis. **a**, Mass spectrum of condensation products between ubiquitin and glycerol, and ubiquitin and Tris. Expected mass for ubiquitin condensation with glycerol = 8,639 Da; found mass = 8,637 Da. Expected mass for ubiquitin condensation with Tris = 8,668 Da; found mass = 8,666 Da. This experiment was repeated twice with similar results. **b**, Chemical structures of Tris and glycerol. **c**, Discharge activity towards Tris and glycerol for all of the tested MYCBP2(cat) cysteine-to-serine mutants (selected mutants shown in Fig. 2c). The C4506S mutation abolishes discharge activity, but because C4506S resides within a Cys-X-X-Cys Zn-binding motif, this was assumed to be a structural defect rather than a catalytic defect. The C4561S mutation undergoes aberrant thioester adduct formation; this may be because the C4561S mutation (also in a structurally important Cys-X-X-Cys Zn-binding motif) unfolds the protein and liberates Cys

residues, which would otherwise be occupied as Zn ligands. These experiments were repeated twice with similar results. **d**, Coomassie stain of the thioester-ester trapping assay with GST-MYCBP2(cat). After the in-gel fluorescence scan, as shown in Fig. 2e, the gel was Coomassie stained. The experiment was repeated at least three times with similar results. **e**, The RCR E3 ligase activity is dependent on both C4520 and C4572. The combination of an inactive C4520S mutant with an inactive C4572A mutant did not restore activity; hence there appears to be cis-cooperation between these two residues (*elevated concentrations of E3 mutants). This experiment was repeated twice with similar results. **f**, Furthermore, consistent with cis-cooperation, SEC-MALS data for untagged MYCBP2(cat) were consistent with a monodisperse species with a calculated molecular weight of 30.06 ± 6.00 kDa (theoretical molecular weight of MYCBP2(cat) monomer = 30.08 kDa). The experiment was repeated twice with similar results.

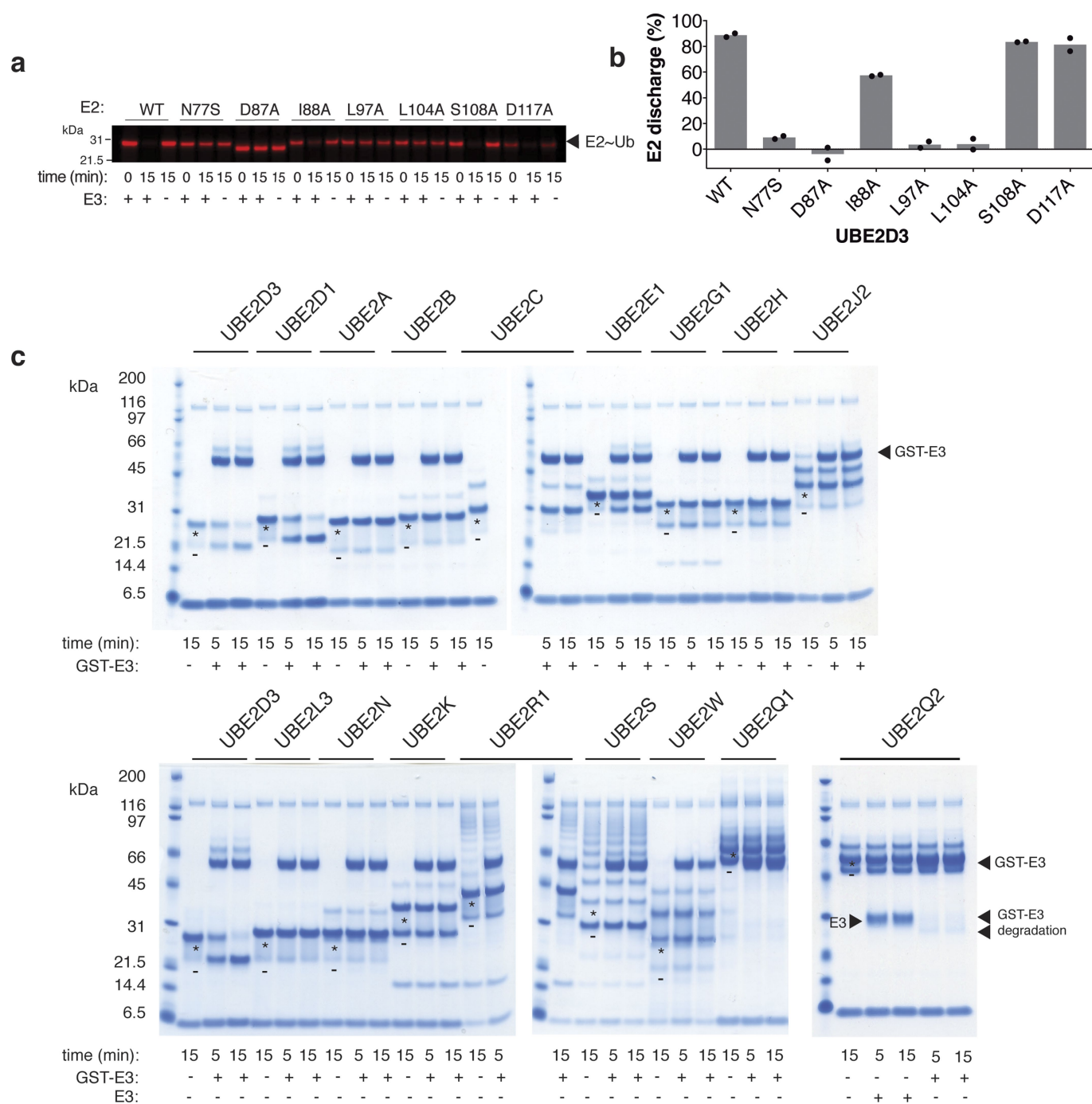


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | MYCBP2 has serine/threonine ubiquitin esterification activity with a preference for threonine.

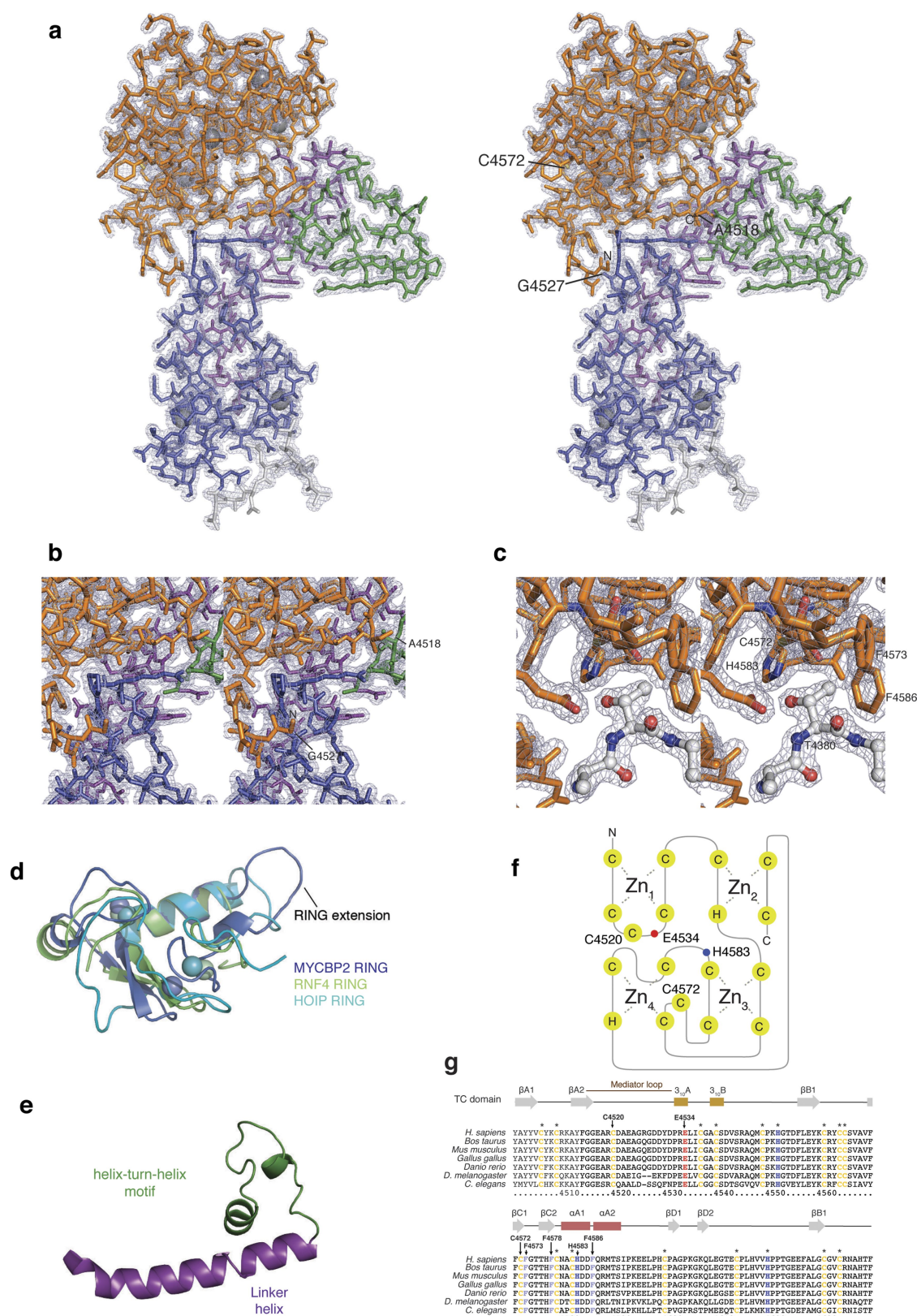
a, HPLC chromatogram of discharge reaction of wild-type MYCBP2(cat) onto threonine (50 mM). Note that esterified threonine with a free amino terminus can undergo O–N acyl transfer, forming a peptide-linked species. **b**, Integrated single-quadrupole electrospray-ionization mass spectrum of the entire peak highlighted in the above chromatogram. Inset shows the deconvoluted mass spectrum (as shown in Fig. 3b). Expected mass of Thr–Ub = 8,666 Da; found mass = 8,664 Da. **c**, HPLC chromatogram of MYCBP2(cat)(C4520S) discharge reaction in the presence of threonine (50 mM). **d**, Integrated single-quadrupole electrospray-ionization mass spectrum of the entire peak highlighted in the above chromatogram. Inset shows the deconvoluted mass spectrum (as shown in Fig. 3b). Expected mass of unmodified ubiquitin = 8,565 Da; found mass = 8,563 Da. All of the above experiments were repeated three times with similar results. **e**, Deconvoluted mass spectra for ubiquitin species in the presence of amino acid (50 mM). The intensities of the ubiquitin reactant and product are reflective of their relative abundance. Observed molecular weight of ubiquitinated serine (Ub–Ser) = 8,650 Da; theoretical molecular weight = 8,652 Da. The observed mass at 8,591 Da corresponds to a side product that is only observed after extended incubation. Ubiquitinated lysine (Ub–Lys) observed molecular weight = 8,691 Da; theoretical molecular weight = 8,693 Da. Assuming exponential ubiquitin consumption, $t_{1/2}$ is around 5 min for threonine. For serine, $t_{1/2}$ is tenfold slower. Lysine ubiquitination is E3-independent as a similar degree of modification is observed in the absence of E3. The experiment was repeated twice with similar results.

f, Coomassie stain of threonine gel presented in Fig. 3c. Also shown is the deconvoluted mass spectrum representative of all ubiquitin species at the 60 min time point. Observed mass of Cy3B–Ub modified threonine peptide = 10,033 Da; theoretical mass = 10,036 Da. **g**, Coomassie stain of the serine gel presented in Fig. 3c. Also shown is the deconvoluted mass spectrum representative of all ubiquitin species at the 60-min time point. Observed mass of Cy3B–Ub = 9,534 Da; theoretical mass = 9,537 Da. Observed mass of Cy3B–Ub modified serine peptide = 10,019 Da; theoretical mass = 10,022 Da. **h**, Coomassie stain of lysine gels, in the presence and absence of E3, presented in Fig. 3c. Inefficient modification of the lysine peptide is observed, which is moderately enhanced in the absence of E3. Experiments shown in **f–h** were repeated more than three times. **i**, Top, observed rate constant (0.024 min^{-1}) for MYCBP2(cat)-threonine-mediated single-turnover E2–Ub discharge, determined by in-gel fluorescence of Cy3B-labelled ubiquitin. The E2 was UBE2D3 and the substrate was threonine (50 mM) ($n = 3$). Bottom, representative replicate gel used for quantification. **j**, Top, observed rate constant for UBE3C-lysine mediated single-turnover E2–Ub discharge was too slow to measure. The E2 was UBE2L3 and the substrate was lysine (50 mM) ($n = 2$). Bottom, representative replicate gel used for quantification. **k**, Top, observed rate constant (0.52 min^{-1}) for HHARI-lysine mediated single-turnover E2–Ub discharge. The E2 was UBE2L3 and the substrate was lysine (50 mM) ($n = 2$). The major component of this rate is attributable to autoubiquitination of lysine residues within HHARI because when lysine is withheld, k_{obs} HHARI-mediated E2–Ub discharge is 0.39 min^{-1} and this is only partially outcompeted by the addition of lysine (n corresponds to the number of biologically independent experiments).



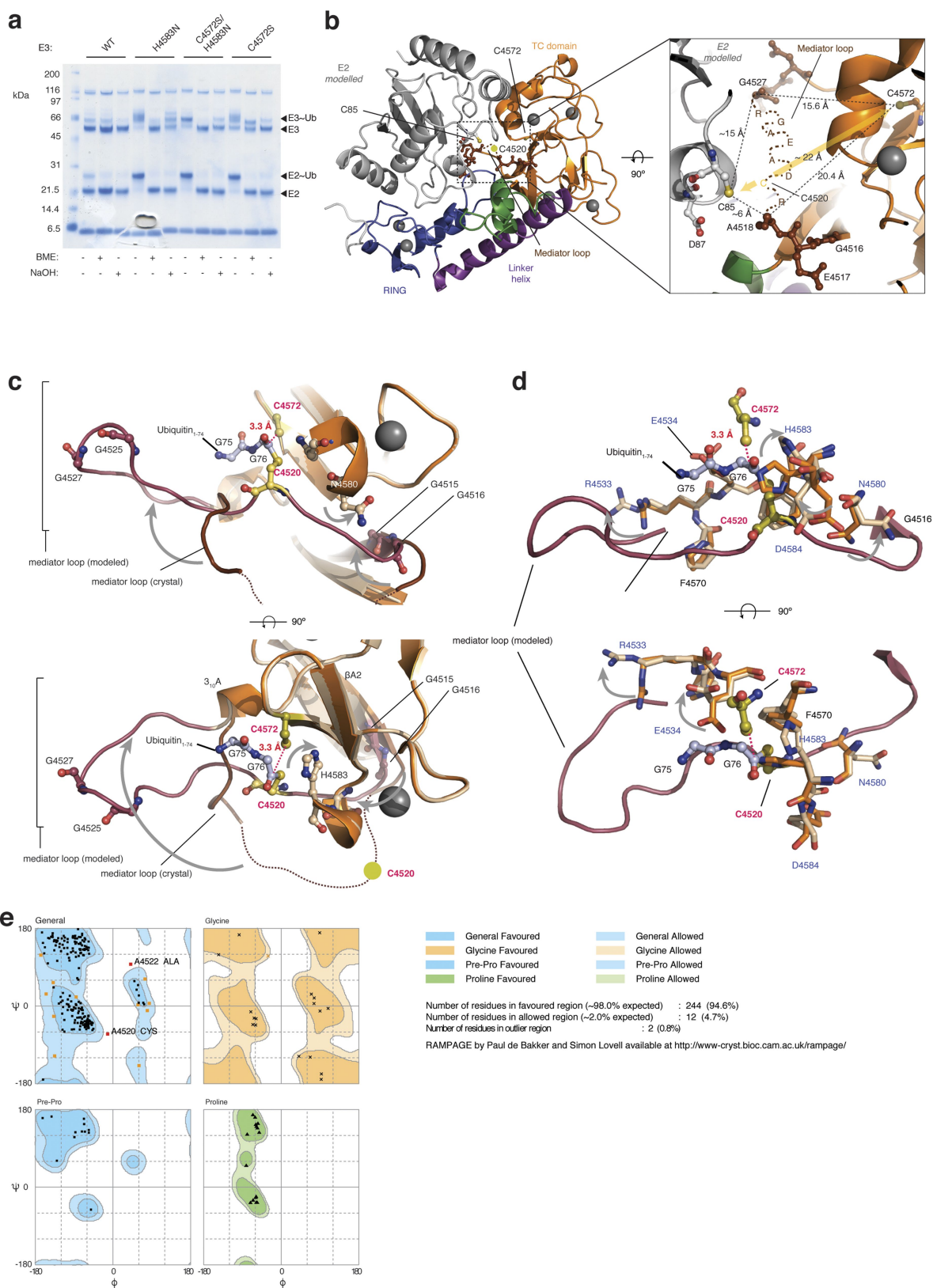
Extended Data Fig. 6 | E2 requirements of MYCBP2. **a**, To establish whether RCR E3 ligase activity occurs exclusively via the proposed E3–Ub thioester intermediates, or alternatively, mediates direct transfer of ubiquitin from E2–Ub (characteristic of RING E3s), we tested RCR E3 activity with a number of UBE2D3 mutants (N77S, D87A, I88A, L97A, L104A, S108A and D117A) that can be diagnostic for these two scenarios⁵. Single-turnover E2–Ub discharge assays employing Cy3B-labelled ubiquitin demonstrate that MYCBP2 has E2 requirements that are consistent with neither a HECT/RBR nor a RING mechanism. The N77 and D117 mutations alter E2 amino acids involved in pK_a suppression of the acceptor nucleophile and are required for RING activity^{35,36}. Additionally, a characteristic of RING activity is the adoption of a ‘closed’ E2–Ub conformation that involves E2 residues D87, I88, L97, L104 and S108. Unlike the RING requirements, UBE2D3, S108 and D117 were dispensable for E2–E3 transthiolation activity. The I88A mutant had reduced activity, the N77S mutant had strongly impaired activity, whereas the D87A, L97A and L104A mutants had negligible activity. Furthermore, RBR E3 activity is permissive to the L104A mutant¹⁸. Thus, based on our current understanding of the E2 requirements of these E3

classes, MYCBP2 has E2 requirements that are consistent with neither a HECT/RBR-like mechanism nor a RING-like mechanism. However, we cannot formally exclude the possibility that MYCBP2_{cat} induces a closed E2–Ub conformation, characteristic of RING E3s, as it does not contain a prohibitive RING-domain loop insertion¹⁹. **b**, Quantification of the different E2 mutant activities. Mean of percentage E2–Ub discharge. $n=2$ biologically independent experiments. **c**, Seventeen E2s were tested for threonine discharge activity with GST–MYCBP2(cat). UBE2D1, UBE2D3 and UBE2E1 were the only E2s that demonstrated detectable activity. –, position of unmodified E2; *ubiquitin-charged E2. Unexpectedly, the HECT/RBR-specific E2 UBE2L3 shows negligible activity with MYCBP2. Certain E2s undergo E3-independent polyubiquitin chain formation and/or autoubiquitination. In the presence of UBE2Q2, GST–MYCBP2 undergoes minor degradation resulting in the appearance of two lower-molecular-weight species. Consequently, we also carried out the assay with untagged MYCBP2 that did not undergo degradation; this produced similar results, showing that UBE2Q2 does not support MYCBP2 activity. The experiment was repeated twice with similar results.



Extended Data Fig. 7 | Structural comparison and representative stereo views of the crystallographic model of MYCBP2(cat). **a**, Wide-field view. Regions are distinguished by colour in the stick representation: RING domain (blue), linker helix (purple), helix–turn–helix motif (green) and tandem-cysteine domain (orange). The mesh represents the experimental $2|F_{\text{obs}}| - |F_{\text{calc}}|$ electron density map contoured at 1.5σ . C4572 is the downstream catalytic cysteine residue in the esterification site. The mediator loop region is formed between A4518 and G4527 and is disordered in the structure. **b**, Close up of the mediator loop region. **c**, Close up of the esterification site. T4380 motif from the symmetry-related molecule (T4380(sym)) is shown and represented in grey ball and

stick. **d**, Superposition of MYCBP2 RING domain with the RING domain from the canonical RING E3 ligase RNF4¹⁶, and from the RBR E3 ligase HOIP¹⁸. **e**, The linker helix and helix–turn–helix motif that connect the RING domain to the tandem cysteine domain. **f**, Diagram depicting the Zn coordination network for the tandem cysteine domain. Catalytic residues (numbered) are distributed throughout the tandem cysteine polypeptide. **g**, The tandem cysteine domain that confers threonine specificity is present in all MYCBP2 orthologues. All residues shown to be required for threonine esterification activity are conserved. Asterisks correspond to Zn-binding residues, grey arrows correspond to β -strands, gold rectangles correspond to 3_{10} -helices, and the red cylinder corresponds to an α -helix.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Modelling of E2-MYCBP2(cat) complex.

a, Ubiquitin adduct formation for catalytic mutants of GST-tagged MYCBP2(cat). The H4583N mutant undergoes near-quantitative ubiquitin-adduct formation. The adduct is largely removed after thiol treatment, indicating that ubiquitin is linked to the E3 via a thioester. The diffuse nature of the upper band might be due to the presence of a trapped thioester-linked ubiquitin on C4520, and C4572, as the H4583N mutation prevents substrate deprotonation. The C4572S/H4583N double mutant forms only a single ubiquitin adduct that is thioester-linked, presumably to the C4520 residue. This indicates that formation of the engineered ester-linked adduct on a mutated S4572 residue is dependent on the presence of a general base. C4520 does not appear to have a base in its proximity, hence its activity could be due to its intrinsic pK_a which results in it being nucleophilic at physiological pH in the absence of a general base. This could explain why we failed to produce an engineered ester adduct on a mutant S4520 residue, as serine is fully protonated at physiological pH. The experiment was repeated twice with similar results. **b**, Superposition of the RING domain from the RBR E3 ligase HOIP in complex with E2 (PDB ID: 5EDV; ubiquitin linked to E2 has been omitted owing to a direct clash with the tandem cysteine domain¹⁸) allows modelling of the E2 into our structure (grey cartoon representation). The catalytic C85

residue in E2 (mutated in silico from Lys to Cys¹⁸) is proximal to C4520, which undergoes transthioylation with E2–Ub. Right, a top-down close up of the mediator loop region. The eight missing residues that form the mediator loop are shown schematically in brown text. **c**, Model of the proposed ubiquitin relay intermediate as shown in Fig. 5e but from an alternative perspective. In the experimental structure, tandem-cysteine-domain residues are shown in orange and mediator-loop residues are in dark brown. In the model, tandem-cysteine-domain residues are in light orange and mediator-loop residues are in mauve. The modelled E2, based on the superposition in **d**, is in grey cartoon. Essential cysteines C4520 and C4572 are in yellow and coloured by atom type. Ubiquitin residues G75–G76 are in blue ball-and-stick representation and are coloured by atom type. Gly residues in the mediator loop that are likely to be important for loop mobility are displayed in mauve ball and stick and coloured by atom type. N4570 and H4583 side chains have been rotated by the specified angles to relieve steric clash. **d**, As in **c**, but amino acid side chains that have been flipped to relieve steric clash with the modelled mediator loop are labelled in blue. **e**, All phi and psi angles in the modelled structure fall within accepted values as determined by Ramachandran analysis with the RAMPAGE server⁴¹.

Extended Data Table 1 | Data collection and refinement statistics

| | MYCBP2 ₄₃₇₈₄₆₄₀ |
|---------------------------------------|----------------------------|
| Data collection | |
| Space group | P 6 ₁ |
| Cell dimensions | |
| a, b, c (Å) | 82.58, 82.58, 103.30 |
| α , γ , β (°) | 90.00, 90.00, 120.00 |
| Resolution (Å) | 40-1.75 (1.78-1.75) |
| R _{merge} | 0.051 (0.404) |
| I / σ I | 18.3 (2.4) |
| Completeness (%) | 99.3 (93.3) |
| Redundancy | 6.9 (3.0) |
| Refinement | |
| Resolution (Å) | 40.0-1.75 |
| No. reflections | 40007 |
| R _{work} / R _{free} | 0.172/0.196 |
| No. atoms | |
| Protein | 1988 |
| Ligand/ion | 6 |
| Water | 257 |
| B-factors | |
| Protein | 28.2 |
| Ligand/ion | 26.6 |
| Water | 35.5 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.019 |
| Bond angles (°) | 1.9 |

Data were collected from a single crystal. Values in parentheses are for highest-resolution shell.

Extended Data Table 2 | Structural alignments of MYCBP2 tandem cysteine domain

[illegible]

Representative tandem-cysteine-domain sequences (130) from various taxa were isolated and aligned as described above. In virtually all animals, a single MYCBP2 orthologue is the only tandem-cysteine-domain-containing gene. There is a relatively large subfamily of sequences found in ciliates and a few other protists. Typically, each of these organisms contains multiple genes of this family. These proteins are also shorter than the animal orthologues. The alignment was rendered with Boxshade. Residues that are invariant in 50% of sequences or conservatively substituted are shown on black and grey backgrounds, respectively. Finally, Cys and His residues involved in Zn coordination are indicated in blue, while the Cys and His residues involved in the ubiquitin relay are shown in red.

Structural basis for ATP-dependent chromatin remodelling by the INO80 complex

Sebastian Eustermann^{1,2,5}, Kevin Schall^{1,2,5}, Dirk Kostrewa^{1,2}, Kristina Lakomek^{1,2}, Mike Strauss³, Manuela Moldt^{1,2} & Karl-Peter Hopfner^{1,2,4*}

In the eukaryotic nucleus, DNA is packaged in the form of nucleosomes, each of which comprises about 147 base pairs of DNA wrapped around a histone protein octamer. The position and histone composition of nucleosomes is governed by ATP-dependent chromatin remodellers^{1–3} such as the 15-subunit INO80 complex⁴. INO80 regulates gene expression, DNA repair and replication by sliding nucleosomes, the exchange of histone H2A.Z with H2A, and the positioning of +1 and –1 nucleosomes at promoter DNA^{5–8}. The structures and mechanisms of these remodelling reactions are currently unknown. Here we report the cryo-electron microscopy structure of the evolutionarily conserved core of the INO80 complex from the fungus *Chaetomium thermophilum* bound to a nucleosome, at a global resolution of 4.3 Å and with major parts at 3.7 Å. The INO80 core cradles one entire gyre of the nucleosome through multivalent DNA and histone contacts. An Rvb1/Rvb2 AAA⁺ ATPase heterohexamer is an assembly scaffold for the complex and acts as a ‘stator’ for the motor and nucleosome-gripping subunits. The Swi2/Snf2 ATPase motor binds to nucleosomal DNA at superhelical location –6, unwraps approximately 15 base pairs, disrupts the H2A–DNA contacts and is poised to pump entry DNA into the nucleosome. Arp5 and Ies6 bind superhelical locations –2 and –3 to act as a counter grip for the motor, on the other side of the H2A–H2B dimer. The Arp5 insertion domain forms a grappler element that binds the nucleosome dyad, connects the Arp5 actin-fold and entry DNA over a distance of about 90 Å and packs against histone H2A–H2B near the ‘acidic patch’. Our structure together with biochemical data⁸ suggests a unified mechanism for nucleosome sliding and histone editing by INO80. The motor is part of a macromolecular ratchet, persistently pumping entry DNA across the H2A–H2B dimer against the Arp5 grip until a large nucleosome translocation step occurs. The transient exposure of H2A–H2B by motor activity as well as differential recognition of H2A.Z and H2A may regulate histone exchange.

Remodellers are grouped into INO80, SWI/SNF, CHD and ISWI families that collectively shape the nucleosome landscape on chromosomal DNA^{7,9}. Although there might be fundamental differences in how remodellers slide, evict and edit nucleosomes^{1–3}, it has been suggested that a common ATP-dependent DNA translocation of the motor domains underlies these distinct reactions³. Recent studies have revealed how the Snf2 motor domain¹⁰ and Chd1 family proteins^{11,12} interact with the nucleosome, but there is currently limited understanding of how stepwise DNA translocation results in its various large-scale reconfigurations. INO80 and the related SWR1 complex are large (megadalton) modular complexes^{13–15} that carry out intricate editing reactions. SWR1 incorporates H2A.Z¹⁶ whereas INO80 has been shown to exchange H2A.Z with H2A^{5,8}. H2A.Z is a H2A variant found at promoter and enhancer elements and has important regulatory functions¹⁷. INO80 also slides nucleosomes and positions the –1 and +1 nucleosomes of genic arrays that flank nucleosome-depleted promoter regions^{6–8}. However, even nucleosome sliding

requires extensive inter-subunit coordination^{18,19} and a clear mechanistic framework explaining these activities is currently not available. Biochemical evidence indicates that INO80 translocates and loops DNA at the H2A–H2B interface⁸, suggesting that sliding and editing may be facets of a common, complex chemo-mechanical reaction.

To provide a structural mechanism for nucleosome recognition and remodelling by INO80, we performed cryo-electron microscopy (cryo-EM) analysis of an evolutionarily conserved, recombinant 11-subunit INO80 complex from *Chaetomium thermophilum* bound to a nucleosome (Fig. 1a–c). Our complex comprises the subunits conserved from yeast to man: the main ATPase Ino80 (INO80 denotes the whole complex; Ino80 refers to the catalytic subunit), actin and actin-related proteins Arp4, Arp5 and Arp8, Ino80 subunits Ies2, Ies4 and Ies6, Taf14 and the AAA⁺ ATPases Rvb1 and Rvb2. It lacks the evolutionarily less conserved subunits—which, in yeast INO80, are Ies1, Ies3, Ies5 and Nhp10—and the N-terminal part of Ino80 to which these subunits bind. Biochemical analysis shows a stoichiometric complex that stably binds and remodels nucleosomes (Extended Data Fig. 1), consistent with the activities of similar human^{14,18} and *Saccharomyces cerevisiae*¹³ INO80 complexes. The nucleosome was assembled from human histones H2A, H2B, H3, H4 and a Widom 601 sequence with 50 bp (base pairs) of extranucleosomal DNA that matches the footprint identified for the entire *S. cerevisiae* INO80⁸.

Cryo-electron microscopy and single-particle reconstruction resulted in a map with a global resolution of 4.3 Å and did not require crosslinking or the addition of nucleotides (Extended Data Figs. 2, 3 and Extended Data Table 1). The map reveals how a 590-kDa core module of INO80 (denoted INO80^{core}) comprising Ino80, Arp5, Ies6, Ies2 and Rvb1/Rvb2 recognizes and remodels the 200-kDa nucleosome core particle (NCP) (Fig. 1c). Focused refinement resulted in a 3.7 Å map of the Rvb1/Rvb2–Arp5–Ies2–Ies6–Ino80 subcomplex (Extended Data Figs. 2, 3). We built de novo atomic models for the ATP-bound Arp5 actin-fold (denoted Arp5^{core}), Ies2, Ies6 and ADP-bound Rvb1/Rvb2 heterohexamer that incorporated the complete Ino80 ATPase insertion domain (denoted Ino80^{insert}). Pseudo-atomic models for the Ino80 Swi2/Snf2 ATPase domain (termed Ino80^{ATPase}) and the NCP were generated by flexible fitting of crystal structures and homology models (Fig. 1c). DNA visibly protrudes from the NCP and a 20 Å cryo-EM map, obtained from extensive 3D classification, indicates extra-nucleosomal binding of the 200-kDa Arp8 module (actin, Arp4, Arp8, Taf14 and Ies4) (Fig. 1b), consistent with genome-wide promoter DNA binding of Arp8 proximal to the +1 nucleosome in vivo²⁰. However, the Arp8 module proved to be either unstable or too heterogeneous in orientation to yield a high-resolution reconstruction at this stage.

INO80^{core} embraces one entire gyre of the nucleosome and binds in a multivalent fashion to nucleosomal DNA and histones (Fig. 1c). The overall mode of NCP recognition of INO80^{core} closely matches the hydroxyl radical footprints of full *S. cerevisiae* INO80⁸. The two main DNA contacts are to superhelical location (SHL) –6 by the Ino80 ATPase motor⁸, and to SHL –2 and SHL –3 by Arp5 and Ies6.

¹Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. ²Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany. ³Max Planck Institute of Biochemistry, Martinsried, Germany. ⁴Center for Integrated Protein Science, Munich, Germany. ⁵These authors contributed equally: Sebastian Eustermann, Kevin Schall. *e-mail: hopfner@genzentrum.lmu.de

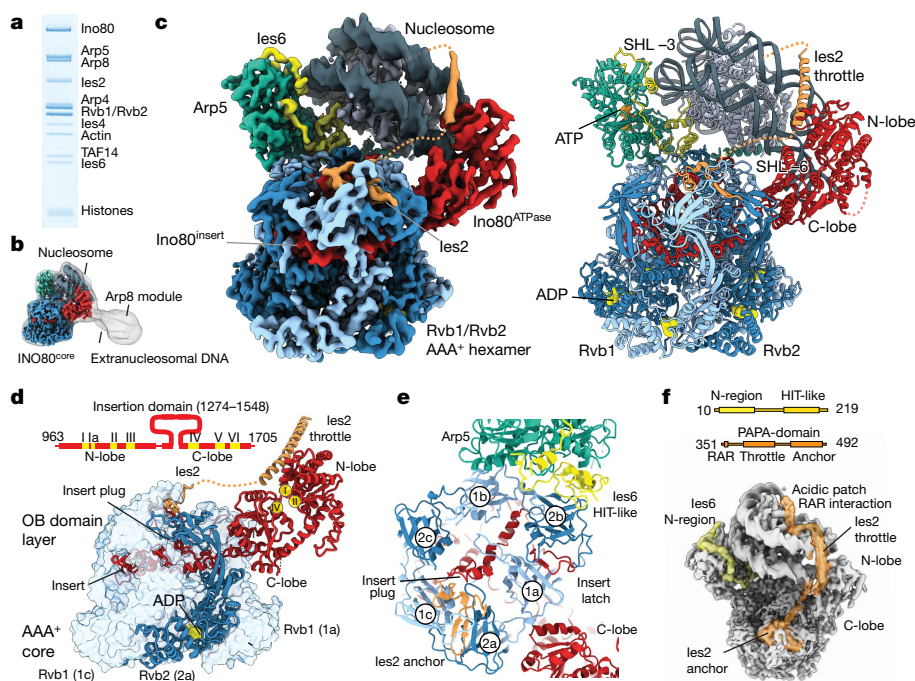


Fig. 1 | Structure of the INO80^{core}-nucleosome complex. **a**, Gel electrophoresis analysis of the purified recombinant *C. thermophilum* INO80 complex bound to a nucleosome. **b**, Low-resolution cryo-EM map showing extra density for the Arp8 module and extranucleosomal DNA. The high-resolution structure of INO80^{core}, shown in **c** and **d**, is superimposed. **c**, Left, 4.3 Å resolution cryo-EM map reveals the architecture of the nucleosome-remodelling core of INO80. Grey, nucleosome; red, Ino80^{ATPase}; orange, Ies2; green, Arp5; yellow, Ies6; light blue, three Rvb1 subunits; dark blue, three Rvb2 subunits. Right, protein models obtained from interpretation of the cryo-EM map showing how the INO80^{core} binds the NCP. ADP and ATP molecules are indicated. The Rvb1/Rvb2 hexamer is assembled from three Rvb1/Rvb2 pairs (denoted 1a, 1b and 1c, and 2a, 2b and 2c; see **e**) and organizes the nucleosome-binding

elements Arp5-Ies6, Ies2 and Ino80^{ATPase}. **d**, Schematic of the Ino80^{ATPase} showing the location of conserved helicase motifs (I-VI) and the insert characteristic of the INO80 family. The insert has a wheel-like structure that binds as a client into the chamber of the three-layered Rvb1/Rvb2 hexamer. One Rvb1/Rvb2 subunit is shown as a ribbon, and the others as transparent surfaces. **e**, Details of the interactions of Arp5-Ies6, Ies2 and Ino80^{ATPase} clients at the OB domain layer of Rvb1/Rvb2. Plug and latch of Ino80^{insert} recruit Ies2 and Arp5-Ies6 clients through direct interactions and/or orienting OB domains. **f**, Ies2 and Ies6 are extended proteins with multiple binding sites that functionally link Rvb1/Rvb2 to the nucleosome via Arp5 and Ino80^{ATPase}, respectively. Of note, Ies2 wraps around the nucleosome and binds the distal acidic patch. The domain architectures are shown above the map.

In addition, we observe contacts of Ino80^{ATPase} and Ies2 to SHL 2, of the 325-amino-acid-long Arp5 insertion domain (termed the grappler) to the dyad, and of the grappler, Ies2 and Ies6 to the histone core (see below). Binding of SHL -6 by the Ino80^{ATPase} motor differs from the SHL +2-binding of the Chd1^{11,12} (Extended Data Fig. 4) and Isw1a remodellers²¹, which indicates that these complexes possess distinct remodelling mechanisms. The isolated Snf2 motor bound to SHL +2 but also to SHL +6¹⁰. Therefore, clarification of mechanistic similarities and differences between INO80 and SWI/SNF families require more complete structures of SWI/SNF remodellers.

The Rvb1/Rvb2 AAA⁺ ATPase is a prominent module of INO80 family remodellers and might act as an assembly chaperone²². We previously interpreted a low-resolution negative stain map as harbouring the Rvb1/Rvb2 double-hexamer that forms in solution¹³, but our high-resolution structure now shows a single hexamer in nucleosome-bound INO80, consistent with a recently published structure of apo human INO80^{core}²³. However, with the nucleosome-bound state and the resolution to build atomic models for the clients, we can now reveal how Rvb1/Rvb2 specifically assembles INO80 and that it has a key role in defining the functional arrangement of subunits of INO80 for interaction with the NCP. The C-lobe of Ino80^{ATPase} directly binds Rvb1/Rvb2 and contains the approximately 270 amino acid-long Ino80 insertion domain that adopts a wheel-like structure and sequentially binds to all six Rvb1/Rvb2 protomers in the central cavity (Extended Data Fig. 5). Ino80^{insert} binding induces a marked asymmetry in the oligonucleotide/oligosaccharide-binding (OB) domain ring layer that

in turn induces specific recruitment and positioning of Ino80^{ATPase}, Ies2 and Arp5-Ies6 to grab the nucleosome from opposing sides (Fig. 1c).

Ino80^{insert} does not bind to the individual Rvb1/Rvb2 units via a shared sequence or even a common structural fold, but the interactions are governed by different hydrophobic and/or aromatic elements in a manner that resembles how bona fide chaperones may bind partially folded proteins²⁴. Comparison with unliganded dodecameric Rvb1/Rvb2²⁵ reveals client-induced conformational control (Extended Data Fig. 5), consistent with a 16-fold stimulation of the ATP hydrolysis activity of Rvb1/Rvb2 by Ino80 insertion peptides²². However, the observed post-hydrolysis ADP state suggests that Rvb1/Rvb2 transforms into a more stable functional scaffold once the correct set of clients is assembled. A 'latch' in Ino80^{insert} binds between OB domains 1a and 2b and generates distinct interaction sites for Arp5 and Ies6 (at OB domains 2a and 2b, respectively). Notably, the C-terminal domain of Ies6 resembles a histidine triad (HIT) zinc finger fold that has lost the zinc-binding cluster, revealing how HIT domains can specifically bind Rvb1/Rvb2 in various complexes²⁶. A 'plug' closes the hole in the OB domain layer and directly binds Ies2, which wedges with a β -hairpin between OB domains 2a and 1c (Fig. 1e). Ies2 reaches all the way across from the Rvb1/Rvb2 OB layer via a linker that is flexible but conserved in length, and pins the N-lobe to SHL 2 (Figs. 1f, 2a and Extended Data Fig. 6 b, d). Ies2 wraps around the nucleosome and binds the acidic patch at the distal side of INO80, which links Ino80^{ATPase} to Rvb1/Rvb2 and the nucleosome; this shows how Ies2 acts as a 'throttle' for the remodelling activity of INO80¹⁸.

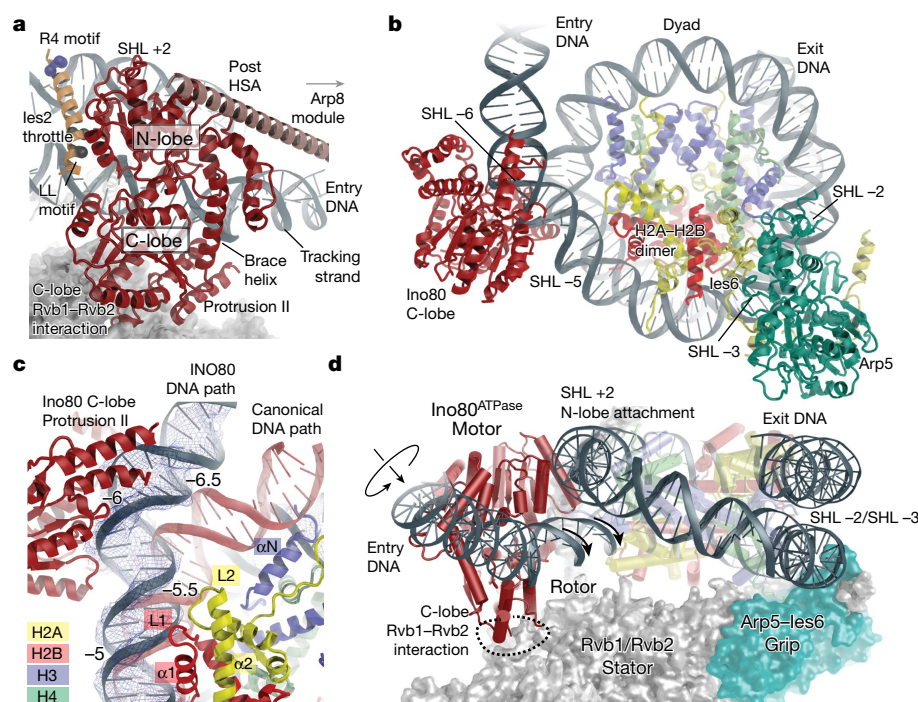


Fig. 2 | Ino80^{ATPase}-nucleosome interaction. **a**, Details of the Ino80^{ATPase}-Ies2 interaction with annotated tracking-stand and notable conserved Ies2 sequence motifs. The post-helicase-SANT-associated (post-HSA) domain (salmon) is provided as a poly-alanine model (Extended Data Fig. 6). **b**, Ino80^{ATPase} and Arp5 bind to opposing sides of the nucleosome, approximately 90 Å apart (for clarity, Rvb1/Rvb2 is not shown). **c**, The binding of Ino80^{ATPase} to exit DNA (blue with superimposed density) unwraps about 15 bp from the nucleosome,

partially exposing H2A at SHL -5.5 and disrupting the H3 interaction at SHL -6.5. The canonical DNA path is shown in red for comparison. The unwrapped DNA is kinked through widening of the minor groove by the C-lobe of Ino80^{ATPase} with its protrusion II element. **d**, Semi-schematic view showing how the Rvb1/Rvb2 hexamer positions the Ino80^{ATPase} motor and Arp5 counter grip on opposite sides of the nucleosome gyre. Rvb1/2 acts as a stator to prevent rotation of the motor with respect to the nucleosome, leading to rotation and translation of entry DNA.

Ino80^{ATPase} is the motor of the remodeller. Conserved Swi2/Snf2 DNA-binding motifs in both the N- and C-lobes engage with double-stranded DNA and the Swi2/Snf2 typical brace helix I reaches across both lobes, stabilizing their mutual orientation (Fig. 2a and Extended Data Fig. 4b). The observed conformation suggests that the motor is poised to bind ATP and to translocate DNA by repetitive cycles of ATP binding and hydrolysis. The binding of Ino80^{ATPase} at SHL -6 unwraps about 15 bp of DNA from the entry site (Fig. 2c). Consequently, DNA contacts to H2A loop 2 (L2) at SHL -5.5 and to H3 helix αN at SHL -6.5 are notably broken, and the H2A-H2B dimer is partially exposed. The full exposure of H2A also requires disruption of the DNA contacts of the loop 1 (L1) and helix α1 of H2A and H2B, which explains why histone exchange additionally requires ATP-driven DNA translocation⁸. The binding of Ino80 ATPase to SHL -6 is accompanied by a widening of the DNA minor groove (Fig. 2c). This finding raises the possibility that the motor domain of INO80 is influenced by DNA shape features, which could be of interest in determining nucleosome positioning at promoter regions⁷.

Swi2/Snf2 proteins translocate DNA by minor groove tracking^{27,28}. The orientation of the Swi2/Snf2 motor at SHL -6 suggests that Ino80^{ATPase} pumps entry DNA into the nucleosome, consistent with the activity of INO80 to centre nucleosomes (Fig. 2d and Extended Data Fig. 1). An important and poorly understood feature of remodellers is how such stepwise translocation of the motor on DNA leads to large-scale reconfiguration of the nucleosome. Building up force on DNA in a processive manner through multiple consecutive steps requires arresting the motor with respect to the nucleosome. The motor of INO80 is fixed by multiple interactions. Ies2 and a secondary DNA-binding site pin the N-lobe to SHL 2. Importantly, the C-lobe is held in place by Rvb1/Rvb2. Rvb1/Rvb2 therefore acts in conjunction with Arp5-Ies6 as a stator, enabling Ino80^{ATPase} to apply force onto the 'rotor' DNA and to pump DNA into the nucleosome. This provides the means of conducting large-scale reconfigurations through multiple translocation steps.

Here we identify Arp5-Ies6 as a major nucleosome recognition module with multiple DNA and histone contacts with both the Arp5 actin fold and the 325-residue-long insertion domain of Arp5 (Arp5^{insert}) that forms a multi-armed grappler (Fig. 3a, b). The C-terminal HIT fold of Ies6 packs in between Rvb1/Rvb2 OB domain 2b and the histone core H2B αC while the conserved N-region of Ies6 wraps around the Arp5 actin fold at the nucleosome proximal DNA side (Fig. 3c, d). Arp5-Ies6 binds about 7–8 bases at SHL -2 and SHL -3, with both Ies6 and a DNA-binding domain (DBD) of the Arp5 actin fold (termed Arp5^{DBD}) (Fig. 3c). The DNA interaction explains the hydroxyl radical footprints of full *S. cerevisiae* INO80 on nucleosomes that showed increased protection of SHL -2 and SHL -3⁸. Of note, Arp5^{DBD} is conserved from yeast to humans (Fig. 3c), and is the structural equivalent of the 'DNase I binding loop' of actin. Mutating conserved DNA-binding arginines/lysines markedly affected nucleosome sliding under conditions in which INO80 still displayed robust ATPase activity (Fig. 3f and Extended Data Fig. 7). Decoupling of ATPase and sliding recapitulates effects seen with Arp5 deletions^{13,29} in *S. cerevisiae* INO80 and human INO180^{core}. We conclude that Arp5-Ies6 couples ATPase activity to nucleosome sliding by gripping DNA and providing an anchor to the histone octamer surface during ratchet translocation steps (see below).

The grappler extends from subdomain 4 of the actin fold of Arp5 and has a multi-armed structure with several notable elements, which we have denoted the 'arm', 'leg', 'foot' and 'bar' (Fig. 3a). Masked 3D classification produced a 4.7 Å map (Fig. 3a) and 4.6 Å map (Extended Data Fig. 6), which together showed that the grappler adopts at least two conformations and enabled us to interpret the topology of its secondary structure with a poly-alanine model. The long N-terminal helix of the Arp5^{insert} forms the bar that, in a closed conformation of the grappler, binds along the nucleosomal dyad and spans between the actin fold of Arp5 and entry DNA at SHL -7.5, over a distance of approximately 90 Å. Importantly, the bar can adopt this binding mode as the entry DNA unwraps from the histone octamer owing to binding of the Ino80

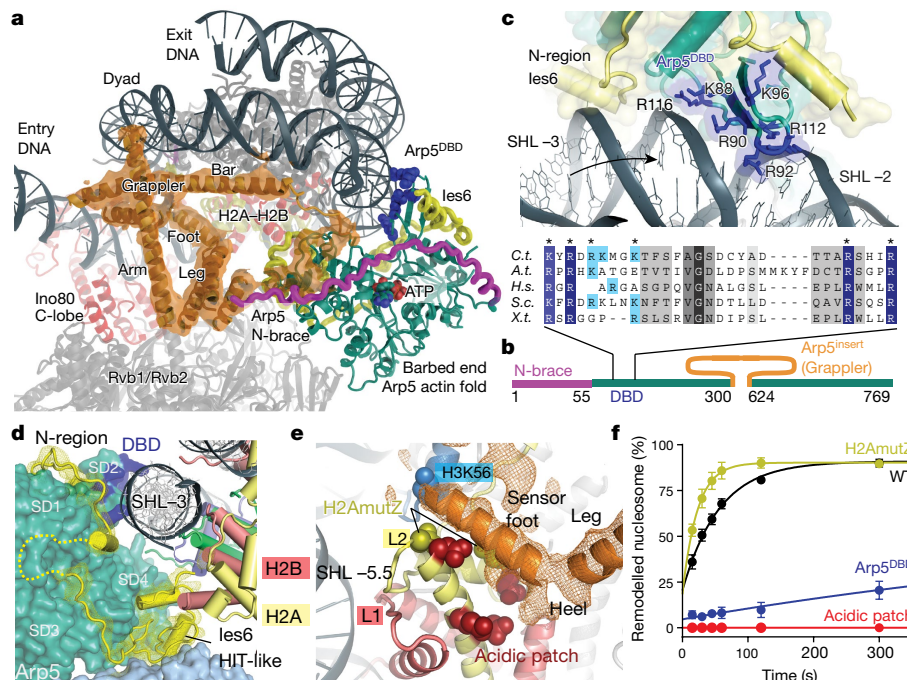


Fig. 3 | Multivalent nucleosome binding by Arp5. **a**, Map at 4.7 Å resolution showing the Arp5 insertion that forms a multi-armed grappler element (orange), along with the actin fold of Arp5 (green, with blue DBD), the Arp5 N-terminal brace (magenta) and Ies6 (yellow). The grappler has multiple DNA and histone contacts and chemo-mechanically connects Arp5, dyad and H2A–H2B. **b**, Schematic of Arp5 domain structure with green actin-fold and highlighted insertions. **c**, Detailed view of the DNA interactions by Ies6 and DBD of Arp5, along with a multiple sequence alignment showing conservation of DNA-binding arginines/lysines in the DBD. Blue: residues mutated for functional analysis (see **e**). *A.t.*, *Arabidopsis thaliana*; *C.t.*, *C. thermophilum*; *H.s.*, *Homo sapiens*; *S.c.*, *S. cerevisiae*; *X.t.*, *Xenopus tropicalis*. **d**, The C-terminal HIT-like domain of Ies6 binds both H2A (yellow) and Rvb1 (light blue), and the N-terminal

region wraps around Arp5^{core}. Actin-fold subdomains SD1–4 are indicated. **e**, Detailed view of the sensor foot and leg of the grappler (orange map and poly-alanine model). The sensor foot binds to the acidic patch of H2A–H2B and to H3 at K56, which suggests it is implicated in controlling histone variant exchange. In **e**, sites mutated (red, acidic patch: E61A, E64A, D72A, D90A; olive, H2A.Z mimic (H2AmutZ): N73L and N89G) for the functional analysis are shown with side chains. **f**, Nucleosome sliding activities of INO80 and histone mutants. H2A.Z-mimicking mutants lead to increased sliding, whereas mutating the H2A acidic patch or Arp5^{DBD} abolishes or strongly reduces sliding under conditions in which INO80 still displays robust ATPase activity. WT, wild-type INO80 with wild-type H2A. Means ± s.d. (*n* = 3) are shown.

ATPase to SHL –6. The arm of the grappler stabilizes the bar at the dyad and connects it to the leg–foot element that packs against the H2A–H2B core at the acidic patch of the histone octamer (Fig. 3e). In an open conformation, the bar is released from the dyad, moves 45° to bind to SHL –1 and blocks the path of the exit DNA (Extended Data Fig. 6). We therefore envision a switch-like behaviour of Arp5 that is sensitive to the path of the entry and exit DNA.

The foot backs H2A opposite L2, as if to stabilize H2A to compensate for the broken DNA contacts that result from the unwrapping of entry DNA. Consequently, the binding of the acidic patch on each side of the nucleosome has an essential role for INO80: the grappler ensures the integrity of the histone octamer where the entry DNA unwraps, and Ies2 binds the acidic patch on the other side of the octamer and acts as a throttle for INO80^{ATPase}. In support of this model, mutating the acidic patch that targets both interactions abrogates nucleosome sliding, although it reduces ATPase rates only moderately (Fig. 3f, Extended Data Fig. 7d, e). Of note, our structure predicts that in a putative dimeric state of INO80²⁹, Ies2 and Arp5 grappler have to compete for the acidic patches on each side of the histone octamer. This might provide asymmetric control of the two INO80 ATPases at SHL –6 and SHL +6 and prevent simultaneous pumping of DNA in opposite directions.

Together with biochemical studies⁸, our structure suggests a unified ratchet-like mechanism for how INO80 slides and possibly edits nucleosomes (Fig. 4). We find that INO80^{core} unwraps entry DNA and grips DNA and histones by multivalent interactions. The motor is positioned to pump DNA into the nucleosome against Arp5–Ies6, which could hold onto DNA until a sufficient force is generated by multiple small steps of the motor. Such groove tracking might create

a DNA loop between the motor and the Arp5–Ies6 counter grip, persistently disrupting the H2A–H2B DNA interface⁸ and thus enabling histone exchange until the amount of DNA pumped propagates across Arp5–Ies6 and the grappler (that is, the ratchet step). As a result, INO80 would move nucleosomes in larger steps (Fig. 4). Step sizes of 10–20 bp

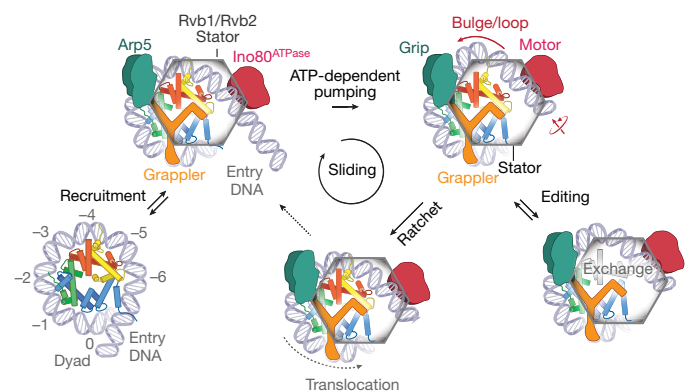


Fig. 4 | Model of INO80 nucleosome remodelling. The unified model integrates our structural data with previous biochemical⁸ data. The functional architecture of INO80 with motor, grip and grappler suggests that processive nucleosome sliding proceeds via a ratchet mechanism. Transient generation of loops between the motor and the grip could expose H2A–H2B for editing. Direct binding of H2A–H2B by the grappler sensor-foot could regulate variant- or modification-specific editing.

are indeed observed^{8,30}. During loop formation, the grappler ensures structural integrity of the octamer by holding onto H2A–H2B, and our structure suggests that its foot also could function as a sensor during editing. The foot binds to H2A at a site at which H2A differs in some amino acid residues from H2A.Z (Extended Data Fig. 7a). Introducing two H2A.Z-mimicking mutations into H2A at this interface increased sliding velocity (Fig. 3f), consistent with the observed faster sliding of H2A.Z nucleosomes by INO80^{8,18}. We observe a direct contact of the ‘toe’ of the sensor-foot with K56 of histone H3. Although controversial, the acetylation of the K56 of H3 has previously been proposed to promote histone variant exchange by INO80 family remodellers³¹ and has a pivotal role in DNA repair and replication as well as in regulating gene expression homeostasis.

In summary, we provide, to our knowledge, the first structural insights into the mechanism by which DNA translocation by a Swi2/Snf2 ATPase of a multisubunit remodeler governs large-scale repositioning and editing reactions on nucleosomes. The motor, stator and multivalent grip of INO80 enable highly processive sliding without release and large-scale reconfigurations such as editing while keeping the remainder of the nucleosome intact. The proposed ratchet mechanism explains DNA loop formation that results in large translocation steps, as well as the means for ATP-dependent H2A.Z → H2A exchange⁸. Thus, our structure visualizes how nucleosome sliding and editing can be achieved by two facets of the same mechano-chemical cycle and how differential regulation might occur. Future studies are needed to address how other modules that are not part of the conserved INO80^{core} function provide an additional layer of regulation (for example, in a promoter-specific manner) and will reveal how the principles discovered for INO80 apply to other remodeler families.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0029-y>.

Received: 18 September 2017; Accepted: 16 February 2018;
Published online 11 April 2018.

- Bartholomew, B. Regulating the chromatin landscape: structural and mechanistic perspectives. *Annu. Rev. Biochem.* **83**, 671–696 (2014).
- Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
- Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).
- Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541–544 (2000).
- Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200–213 (2011).
- Udugama, M., Sabri, A. & Bartholomew, B. The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol. Cell. Biol.* **31**, 662–673 (2011).
- Krietenstein, N. et al. Genomic nucleosome organization reconstituted with pure proteins. *Cell* **167**, 709–722.e712 (2016).
- Brahma, S. et al. INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat. Commun.* **8**, 15616 (2017).
- Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
- Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2–nucleosome structure. *Nature* **544**, 440–445 (2017).
- Sundaramoorthy, R. et al. Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes. *eLife* **6**, e22510 (2017).
- Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome–Chd1 structure and implications for chromatin remodelling. *Nature* **550**, 539–542 (2017).
- Tosi, A. et al. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–1219 (2013).
- Chen, L. et al. Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling. *J. Biol. Chem.* **286**, 11283–11289 (2011).
- Nguyen, V. Q. et al. Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell* **154**, 1220–1231 (2013).
- Mizuguchi, G. et al. ATP-driven exchange of histone H2A.Z variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).
- Albert, I. et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
- Willhoft, O., Bythell-Douglas, R., McCormack, E. A. & Wigley, D. B. Synergy and antagonism in regulation of recombinant human INO80 chromatin remodeling complex. *Nucleic Acids Res.* **44**, 8179–8188 (2016).
- Chen, L., Conaway, R. C. & Conaway, J. W. Multiple modes of regulation of the human INO80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proc. Natl Acad. Sci. USA* **110**, 20497–20502 (2013).
- Yen, K., Vinayachandran, V. & Pugh, B. F. SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell* **154**, 1246–1256 (2013).
- Gangaraju, V. K. & Bartholomew, B. Dependency of ISW1a chromatin remodeling on extranucleosomal DNA. *Mol. Cell. Biol.* **27**, 3217–3225 (2007).
- Zhou, C. Y. et al. Regulation of Rvb1/Rvb2 by a domain within the INO80 chromatin remodeling complex implicates the yeast Rvbs as protein assembly chaperones. *Cell Reports* **19**, 2033–2044 (2017).
- Aramayo, R. J. et al. Cryo-EM structures of the human INO80 chromatin-remodeling complex. *Nat. Struct. Mol. Biol.* **25**, 37–44 (2017).
- Koldewey, P., Horowitz, S. & Bardwell, J. C. A. Chaperone–client interactions: non-specificity engenders multifunctionality. *J. Biol. Chem.* **292**, 12010–12017 (2017).
- Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K. P. Structural basis for dodecameric assembly states and conformational plasticity of the full-length AAA + ATPases Rvb1•Rvb2. *Structure* **23**, 483–495 (2015).
- Rivera-Calzada, A. et al. The structure of the R2TP complex defines a platform for recruiting diverse client proteins to the HSP90 molecular chaperone system. *Structure* **25**, 1145–1152.e4 (2017).
- Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* **16**, 2120–2134 (2002).
- Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
- Willhoft, O. et al. Crosstalk within a functional INO80 complex dimer regulates nucleosome sliding. *eLife* **6**, e25782 (2017).
- Zhou, C. Y. et al. The yeast INO80 complex operates as a tunable DNA length-sensitive switch to regulate nucleosome sliding. *Mol. Cell* **69**, 677–688.e9 (2018).
- Watanabe, S., Radman-Livaja, M., Rando, O. J. & Peterson, C. L. A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science* **340**, 195–199 (2013).

Acknowledgements We thank P. Korber and E. Oberbeckmann for help with biochemical analysis of INO80, E. Hurt for help with *C. thermophilum* genome annotations, D. Berninghausen for help with data collection, and K. Lammens, K. Knoll, R. Byrne, G. Stöhr and G. Witte for discussions and technical help. We thank the LMU Protein Analysis Unit and the MPI of Biochemistry cryo-EM and biophysics core facilities for access and support, and the SuperMUC for computing time. K.-P.H. is supported by the Deutsche Forschungsgemeinschaft (CRC1064, RTG1721), the European Research Council (ERC Advanced Grant ATMMACHINE), the Gottfried-Wilhelm-Leibniz Prize and the Center for Integrated Protein Sciences Munich (CIPSM). S.E. acknowledges an EMBO long-term fellowship and K.S. acknowledges funding by Quantitative Biosciences Munich (QBM).

Reviewer information Nature thanks B. Bartholomew, O. Llorca and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.E. prepared cryo-EM samples and performed structure determination. K.S. prepared nucleosomes, performed the biochemical analysis and participated in structure determination. S.E., D.K. and K.-P.H. build atomic models. K.L. characterized INO80 subcomplexes and participated in early stages of the electron microscopy analysis. M.S. operates the MPI Biochemistry cryo-EM facility, helped with electron microscopy data collection and provided general electron microscopy advice. M.M. helped with recombinant protein production. S.E. and K.-P.H. designed the overall study, analysed the results and wrote the paper with contributions from all other authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0029-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0029-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to K.-P.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

INO80 expression and purification. *C. thermophilum* INO80 subunits were cloned and expressed using the MultiBac technology³². Genes coding for Ino80^{718–1848} with a C-terminal 2 × Flag, Taf14 and Ies4 were each cloned in pACEBac1, Rvb2 and actin in pIDC, Ies2 and Arp8 in pIDS and Ies6, Rvb1, Arp5 and Arp4 in pIDK. Resulting gene cassettes coding for Ino80^{718–1848} 2 × Flag, Rvb1, Rvb2, Arp5, Ies2 and Ies6 were combined in one bacmid, whereas those coding for Ies4, Taf14, Arp8, actin and Arp4 were combined in a separate bacmid. Recombination steps were carried out in *Escherichia coli* XL1-Blue cells (Stratagene) or pirHC cells (Geneva Biotech) under addition of Cre recombinase (NEB). Baculoviruses were generated in *Spodoptera frugiperda* (SF21) insect cells (IPLB-Sf21AE). *Trichoplusia ni* High Five cells (Invitrogen) were co-infected with 1/100 v/v of each baculovirus. Hi5 and SF9 insect cells were purchased from Invitrogen and used for protein production without further authentication. Cells were cultured for 60 h at 27 °C and collected by centrifugation. For complex purification, cells were disrupted in lysis buffer (30 mM HEPES, pH 7.8, 300 mM NaCl, 10% glycerol, 20 μM ZnCl₂, 0.25 mM DTT, 0.28 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamide) and gently sonified. Raw lysate was cleared by centrifugation at 30500g and 4 °C for 30 min. Supernatant was incubated with 4 ml anti-Flag M2 affinity gel (Sigma-Aldrich) for 1 h and washed with 75 ml lysis buffer and 50 ml wash buffer (30 mM HEPES, pH 7.8, 150 mM NaCl, 5% glycerol, 0.5 mM CaCl₂, 20 μM ZnCl₂, 0.25 mM DTT). The complex was eluted by incubation with 8 ml elution buffer (wash buffer supplemented with 0.2 mg/ml Flag peptide) for 20 min at 4 °C. Next, the sample was loaded onto a Mono Q 5/50 GL column (GE Healthcare) and eluted by a gradient of increasing salt, resulting in a highly pure INO80 sample. **Right-angle light scattering measurement.** Molecular weight of apo INO80 was determined by right-angle light scattering. Size-exclusion chromatography (SEC)-coupled static light scattering was performed using an Äkta micro chromatography system equipped with a Superose 6 10/300 Increase column (GE Healthcare) and a right-angle laser static light scattering device and refractive index detector (Malvern/Viscotek). BSA was used to calibrate the system. Evaluation was performed using the OmniSEC software (Malvern/Viscotek).

Purification of mononucleosomes. Canonical human histones and their mutants were purified by a combination of inclusion body purification and ion-exchange chromatography, essentially as previously described^{33,34}. In brief, histones were expressed in *E. coli* BL21 (DE3) cells (Novagen) for 2 h after induction at 37 °C. Cells were disrupted under non-denaturing conditions and inclusion bodies were washed with 1% Triton X-100. Inclusion bodies were resuspended in 7 M guanidinium chloride, dialysed in 8 M urea and histones were purified by cation-exchange chromatography. After refolding under low-salt conditions, anion exchange chromatography was performed as a final purification step. Histones were lyophilized for long-time storage. For octamer assembly, single histones were resuspended in 7 M guanidinium chloride, mixed at 1.2-fold excess of H2A and H2B and dialysed against 2 M NaCl for 16 h. Histone octamers were purified by size-exclusion chromatography using a Superdex 200 16/60 column (GE Healthcare) and were stored in 50% glycerol at –20 °C. We used the Widom 601 DNA³⁵ with 50 or 80 bp extranucleosomal DNA in the ONX orientation³⁶ for reconstituting mononucleosomes. DNA was amplified by PCR, purified using anion-exchange chromatography and concentrated in vacuum. DNA and histone octamer were mixed at a 1.1-fold excess of DNA at 2 M NaCl and sodium chloride concentration was decreased to 50 mM over 17 h at 4 °C. Finally, nucleosomes were purified by anion-exchange chromatography, dialysed to 50 mM NaCl, concentrated to 1 mg/ml and stored at 4 °C.

Purification and vitrification of the INO80–ON50 complex. INO80 and ON50 (nucleosome flanked by 0- and 50-base-pair extranucleosomal DNA) nucleosomes were mixed at a ratio of 2:1 and dialysed to binding buffer (20 mM HEPES, pH 8, 60 mM KCl, 0.5% glycerol, 0.25 mM CaCl₂, 20 μM ZnCl₂, 0.25 mM DTT) for 1 h in Slide-a-lyzer dialysis tubes (Thermo Fisher Scientific). The complex was purified by gel filtration using a Superose 6 3.2/300 column (GE Healthcare) and vitrified at a concentration of 1 mg/ml on Quantifoil R2/1 grids in the presence of 0.05% octyl-β-glucoside using a Leica EM GP (Leica).

Electron microscopy and data collection. The FEI Titan Krios transmission electron microscope was operated at 300 kV using a GIF quantum energy filter (slit width 20 eV) and a Gatan K2 summit direct electron detector. Two datasets of images (dataset I and dataset II) with a defocus ranging from 1.3 to 3.5 μm were collected at a calibrated pixel size of 1.34 Å and 1.06 Å and at a dose rate of 5.63 and 5.96 e[–]/Å²/s, respectively. A total dose of 67.5 and 59.6 e[–]/Å² was recorded over 12 and 10 s with a frame rate of 5 and 4 frames stored per second for dataset I and dataset II, respectively. Data acquisition was carried out using SerialEM³⁷ facilitated by a set of customized scripts that enabled automated execution of low-dose image acquisition, including focus and drift determination as well as beam centring (M.S. et al., manuscript in preparation).

Cryo-EM data processing. Dose-fractionated image stacks were subjected to beam-induced motion correction using MotionCorr³⁸. The first and the last frame were discarded and CTF parameters for each sum of remaining frames determined by CTFFIND4³⁹. Micrographs that exhibited too much drift, too much contamination or abnormal Fourier patterns were discarded. For dataset I (at 1.34 Å/pixel), 1,282 image stacks were chosen for further processing, and for dataset II (at 1.06 Å/pixel) and 3,932 image stacks were chosen for further processing, carried out using MotionCorr³⁸-corrected sums that were filtered according to exposure dose. Particle selection, 2D classification, 3D classification and refinement were performed using RELION⁴⁰ version 2.1.1b, unless stated otherwise. All resolutions that we report here were determined by gold standard Fourier shell correlation 0.143 criterion. B-factors were automatically determined within RELION according to a previously published method⁴¹. Extended Data Fig. 2a, b shows an overview of the cryo-EM processing scheme used for dataset I and dataset II. Two-dimensional class averages (Extended Data Fig. 2d), used as 35 Å low-pass-filtered templates for the initial automated particle picking of dataset I, were calculated from 800 particles that were manually picked from a screening dataset acquired using a FEI Falcon II camera and a FEI Titan Halo transmission electron microscope at 300 kV. Six thousand semi-automatically picked particles from the same dataset were used to generate a 3D ab initio reconstruction in CryoSPARC⁴² (Extended Data Fig. 2c), which served as a 40 Å low-pass-filtered reference for the first round of 3D classification in RELION. 2D and 3D classification (3D classification A1 and 3D classification A2, Extended Data Fig. 2a) identified 18,000 particles corresponding to nucleosome-bound INO80^{core} complexes from 295,000 automatically picked particles. Because we refrained from crosslinking to stabilize complexes during sample and grid preparation, we observed a large number of disassembled complexes at vitrified conditions corresponding to free nucleosomes (class 1 of 3D classification A2, Extended Data Fig. 2a) or apo INO80^{core} complex (class 3 of 3D classification A2, Extended Data Fig. 2a). Severe orientational bias of particles in this dataset prevented meaningful refinement of the apo INO80^{core} complex beyond 8 Å. By contrast, the identified set of 18,000 particles of nucleosome-bound INO80^{core} subjected to RELION refinement and subsequent solvent mask post-processing yielded a cryo-EM map of the nucleosome complex at an overall resolution of 5.8 Å. This map was used as a reference to determine a higher resolution structure using the larger dataset II recorded at higher magnification (1.06 Å/pixel). To improve auto-picking of sparsely populated orientations of the complex, we calculated 2D projections of the experimentally determined 5.8 Å cryo-EM map (Extended Data Fig. 2e). To avoid false positives during particle picking, we applied a 35 Å low-pass filter to the projections before using them as templates and verified the quality of the automated particle picking procedure by visual inspection of the micrographs as well as by diagnostic 2D classifications in RELION. Two hundred and fifty-two thousand particles derived from automated particle picking were subjected to successive rounds of 3D classification (3D classification B1, 3D classification B2 and 3D classification B3, Extended Data Fig. 2b). Notably, an intermediate set of 144,000 particles yielded a cryo-EM map of the INO80^{core} complex at 3.9 Å. Although inspection indicated there was still conformational or compositional heterogeneity within the region of the nucleosome, the particle density and signal-to-noise ratio was sufficiently high to enable movie processing and particle polishing within RELION (using frames 1–30, running averages of 8 frames and a standard deviation of particles of 300 Å). Subsequent refinement of the ‘polished’ particles yielded a 3.7 Å map that allowed de novo atomic model building and real space refinement of Ies2, Ies6, Arp5, Ino80^{insert} and Arp5 (see below). Three-dimensional classification (3D classification B2) yielded a class of 34,000 nucleosome-bound particles. These particles were subjected to RELION refinement and solvent mask post-processing, yielding a cryo-EM map of the complex at an overall resolution of 4.3 Å (Extended Data Fig. 2f). Finally, two classes showing different conformations of the grapple element were obtained by using a third 3D classification (3D classification B3, Extended Data Fig. 2b) in which the Euler angles derived from the previous refinement were kept fixed and a mask of the respective region of the complex was applied. Local resolution estimation and local resolution filtering was performed as implemented in RELION 2.1.1b.

Model building and refinement. As a first stage we performed rigid-body docking in UCSF Chimera⁴³ using available crystal structures of *Xenopus laevis* nucleosome with Widom 601 sequence (RCSB Protein Data Bank (PDB) code: 4R8P), crystal structures of *C. thermophilum* Rvb1 and Rvb2 (PDB codes: 4WW4 and 4FM6) and homology models of *C. thermophilum* Arp5^{core} residues 59–755 (excluding insert residues 306–640) as well as *C. thermophilum* Ino80^{ATPase} residues 964–1705 (excluding insert residues 1274–1548). A homology model of the actin fold of Arp5 was built using SWISS-MODEL⁴⁴ using ATP-bound actin (PDB code: 1NWK) as a template, while I-TASSER⁴⁵ was used to build separate homology models for the N- and C-lobe of Ino80^{ATPase} using multiple high-resolution X-ray structures of related superfamily 2 ATPases as templates. Atomic model building of Ino80 insert (residues 1278–1544), Ies2 (residues 443–478), Ies6 (residues 10–52, 155–213) and

Arp5 (residues 15–107, 111–146, 153–300, 603–769) was performed using the 3.7 Å map of INO80^{core} and a combination of COOT⁴⁶ and Moloc⁴⁷. Model building and refinement was performed iteratively using restrained real-space refinement in PHENIX 1.12. We used restraints for secondary structure, side chain rotamers, Ramachandran and C β restraints, while we restricted the resolution to 3.7 Å during refinement. In the final macrocycle, grouped B-factor refinement for the main chain and side chain was calculated. Statistics of the final refinement and the obtained structures are reported in Extended Data Table 1. The obtained structures were subsequently used for interpretation and model refinement using the 4.3 Å resolution cryo-EM map of the INO80^{core}–nucleosome complex. To model regions with larger conformational deviations such as the nucleosomal DNA, the Ino80^{ATPase} and regions at the nucleosome interface of INO80^{core} we used a combination of flexible fitting and (re)building using a combination of COOT⁴⁶, Moloc⁴⁷ and MDFF⁴⁸. This procedure resulted in reasonable refinement of model into the cryo-EM map (Extended Data Fig. 3c). The properties and limitations of the molecular models of the INO80^{core}–NCP complex are summarized in the following. Flexible fitting of nucleosomal DNA accounts for the large conformational change seen in the region between SHL –5.5 and SHL –7. Although additional unambiguous density corresponding to extranucleosomal DNA protrudes from the INO80^{core}–NCP complex, we did not attempt to build DNA beyond SHL –7 at this stage. The histone core required only minor adjustments. However, we do not observe density for H3 tail residues 37–44 at their canonical binding site above SHL 1. We observed instead unassigned density between the foot element of the grappler and the N-terminal H3 helix α N. Because this density can also originate from grappler element of Arp5 (see below), we refrained from building the H3 tail at this stage. Ino80 residues 964–1274 and 1549–1705 were flexibly fitted into the density and readily connected to the refined model of the insert region described above. The topology of the grappler element was unambiguously assigned to the Arp5 insert residues 300–624. However, model building was largely restricted to a poly-alanine model given the limited resolution of this element in the 4.6 Å and 4.7 Å subclasses (Extended Data Fig. 3e). Similarly, we were able to build a poly-alanine model of Ino80 post-HSA residues 820–855 and Ies2 residues 351–443 that includes the throttle helix bound to nucleosomal DNA (Extended Data Fig. 6).

Electrophoretic-mobility shift assays. Electrophoretic-mobility shift assays were used to monitor the interaction between INO80 and ON50 mononucleosomes. Nucleosomes were labelled at the 5'-end of their extranucleosomal DNA with fluorescein. Nucleosome (15 nM) was incubated with increasing concentrations of INO80 (0, 5, 10, 20 and 40 nM) in electrophoretic mobility shift assay buffer (25 mM HEPES, pH 8, 60 mM KCl, 7% glycerol, 0.25 mM DTT, 2 mM CaCl₂) for 20 min on ice. Samples were analysed at 4 °C by native PAGE on a 3–12% acrylamide BIS-Tris gel (Invitrogen) and visualized using the Typhoon imaging system (GE healthcare).

Nucleosome sliding assays. ON80 (nucleosome flanked by 0- and 80-base-pair extra-nucleosomal DNA) mononucleosomes with 5'-fluorescein-labelled extranucleosomal DNA were used for monitoring the sliding activity of INO80. Nucleosome (150 nM) was incubated with 50 nM INO80 in sliding buffer (25 mM HEPES, pH 8, 60 mM KCl, 7% glycerol, 0.10 mg/ml BSA, 0.25 mM DTT, 2 mM MgCl₂) at 25 °C. The reaction was started on addition of 1 mM ATP and stopped at several time points (15, 30, 45, 60, 120, 300, 500 and 1,200 s) by addition of 0.2 mg/ml lambda DNA (NEB). Nucleosome species were separated by native PAGE on a 3–12% acrylamide BIS-Tris gel (Invitrogen) and visualized using the Typhoon imaging system (GE healthcare). ImageJ was used to quantify gel bands and the fraction of remodelled band was plotted against the reaction time. Data describe a saturation curve and were fitted in Prism (GraphPad) using an exponential equation.

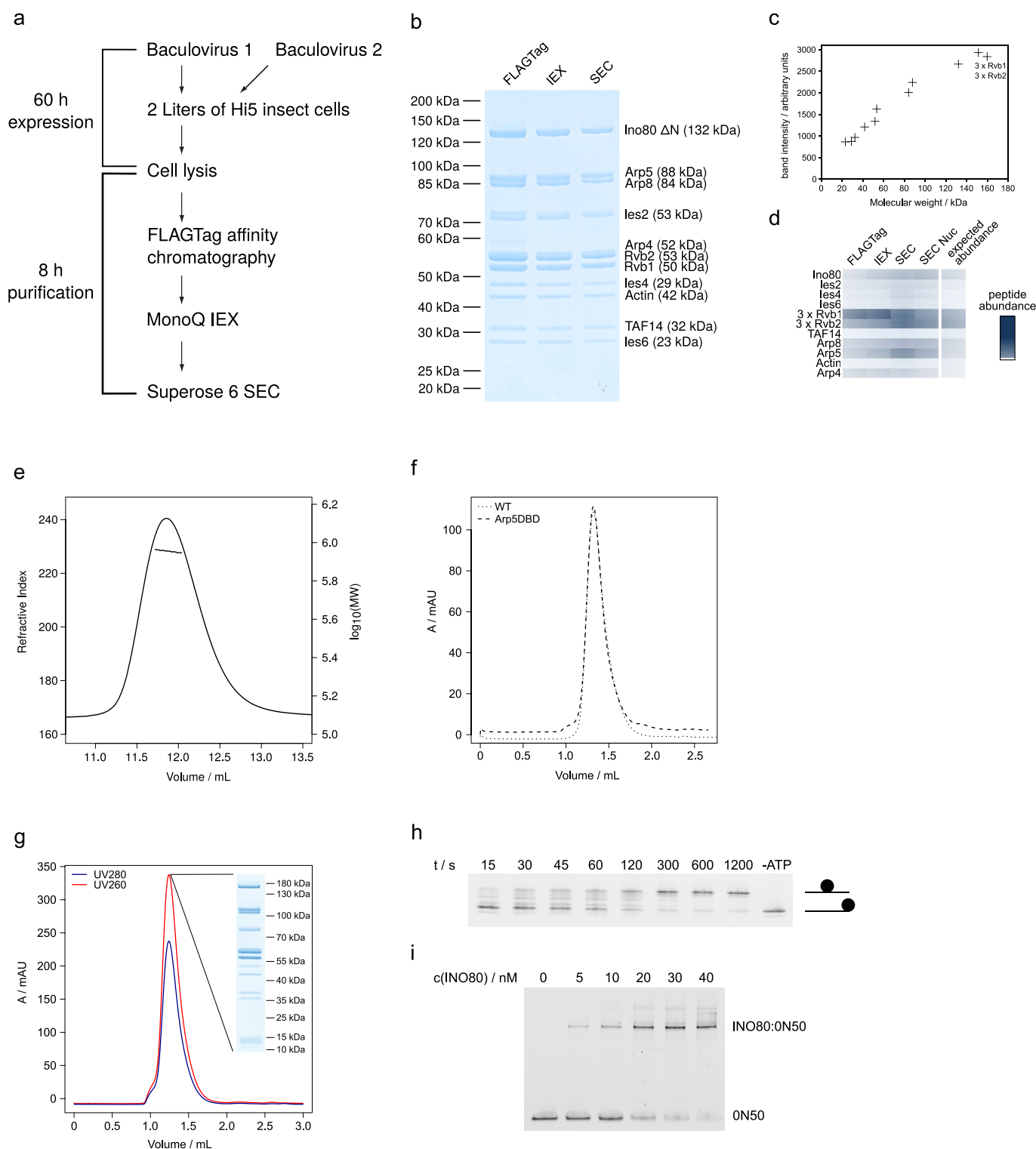
ATPase assays. An ATPase assay coupling ATP hydrolysis to NADH oxidation was used to determine the ATPase rate of INO80. INO80 (30 nM) was incubated

in assay buffer (25 mM HEPES, pH8, 50 mM KCl, 1 mM DTT, 2 mM MgCl₂, 0.1 mg/ml BSA) with 0.5 mM phosphoenolpyruvate, 1 mM ATP, 0.1 mM NADH and 25 U/ml lactate dehydrogenase and pyruvate kinase (Sigma) at 25 °C in a final volume of 50 μ l. NADH concentration was monitored fluorescently over 1 h in non-binding black 384-well plates (Greiner Bio-One) using 340 nm for excitation and an emission of 460 nm with a Tecan Infinite M100 (Tecan). Where indicated, ATPase activity was determined in the presence of 150 nM nucleosome. ATP turnover was calculated using maximal initial linear rates, corrected for a buffer blank.

Figure preparation. Figures were prepared with PyMol (The PyMOL Molecular Graphics System, version 1.8 Schrödinger, LLC), UCSF Chimera⁴³ and UCSF ChimeraX⁴⁹.

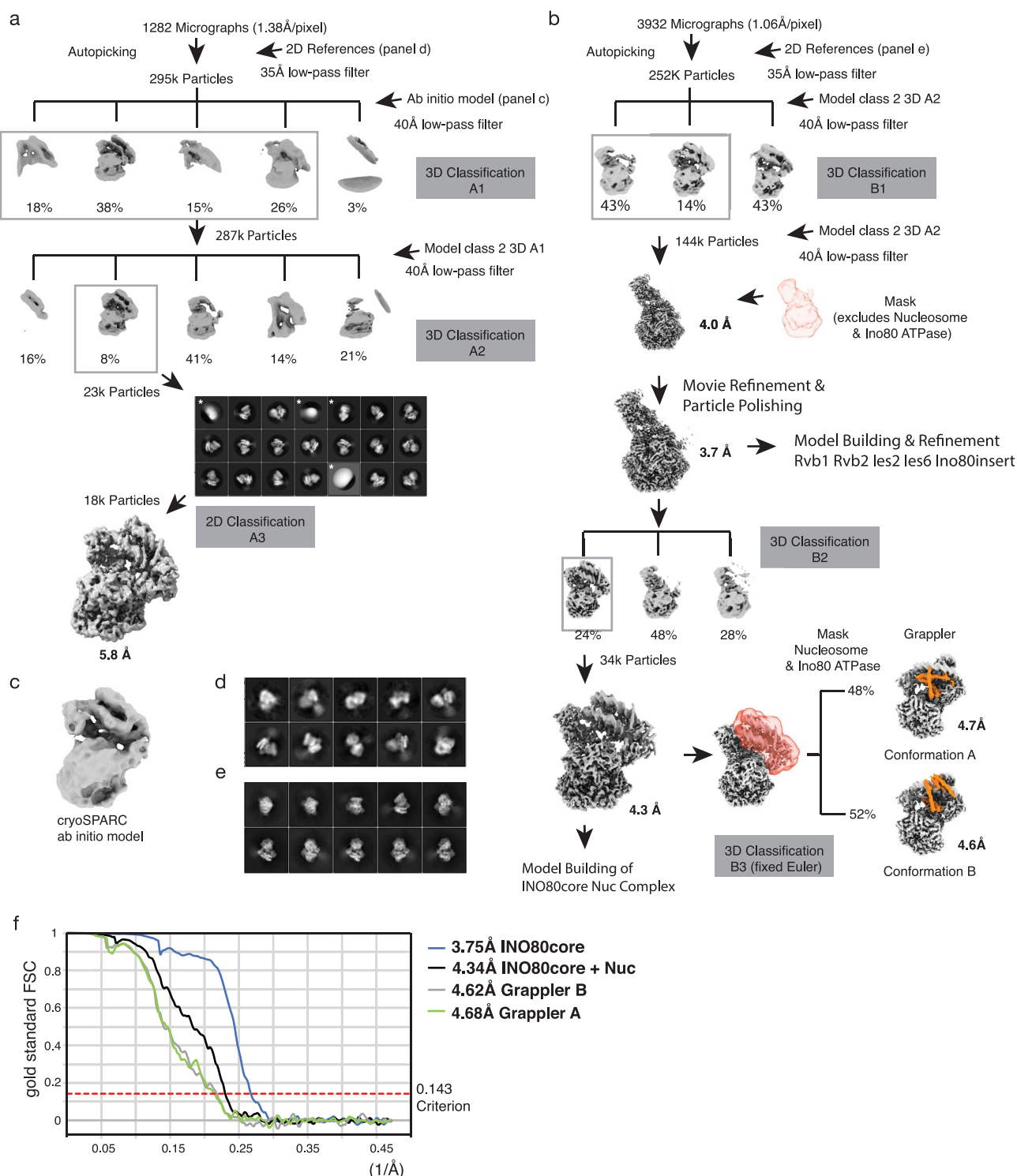
Data availability. The electron density reconstruction and final model have been deposited with the Electron Microscopy Data Base under accession codes EMD-4264, EMD-4277, EMD-4278 and EMD-4280, and with the RCSB Protein Data Bank under accession codes 6HFS and 6FML. Uncropped images of the polyacrylamide gels are shown in Supplementary Fig. 1. All other data are available from the corresponding author upon reasonable request.

32. Fitzgerald, D. J. et al. Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021–1032 (2006).
33. Klinker, H., Haas, C., Harrer, N., Becker, P. B. & Mueller-Planitz, F. Rapid purification of recombinant histones. *PLoS ONE* **9**, e104029 (2014).
34. Dyer, P. N. et al. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
35. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
36. Levendosky, R. F., Sabantsev, A., Deindl, S. & Bowman, G. D. The Chd1 chromatin remodeler shifts hexasomes unidirectionally. *eLife* **5**, e21356 (2016).
37. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
40. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
41. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
42. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
43. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
45. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. Gerber, P. R. & Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9**, 251–268 (1995).
48. Qi, Y. et al. CHARMM-GUI MDFF/xMDFF utilizer for molecular dynamics flexible fitting simulations in various environments. *J. Phys. Chem. B* **121**, 3718–3723 (2017).
49. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2017).
50. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).



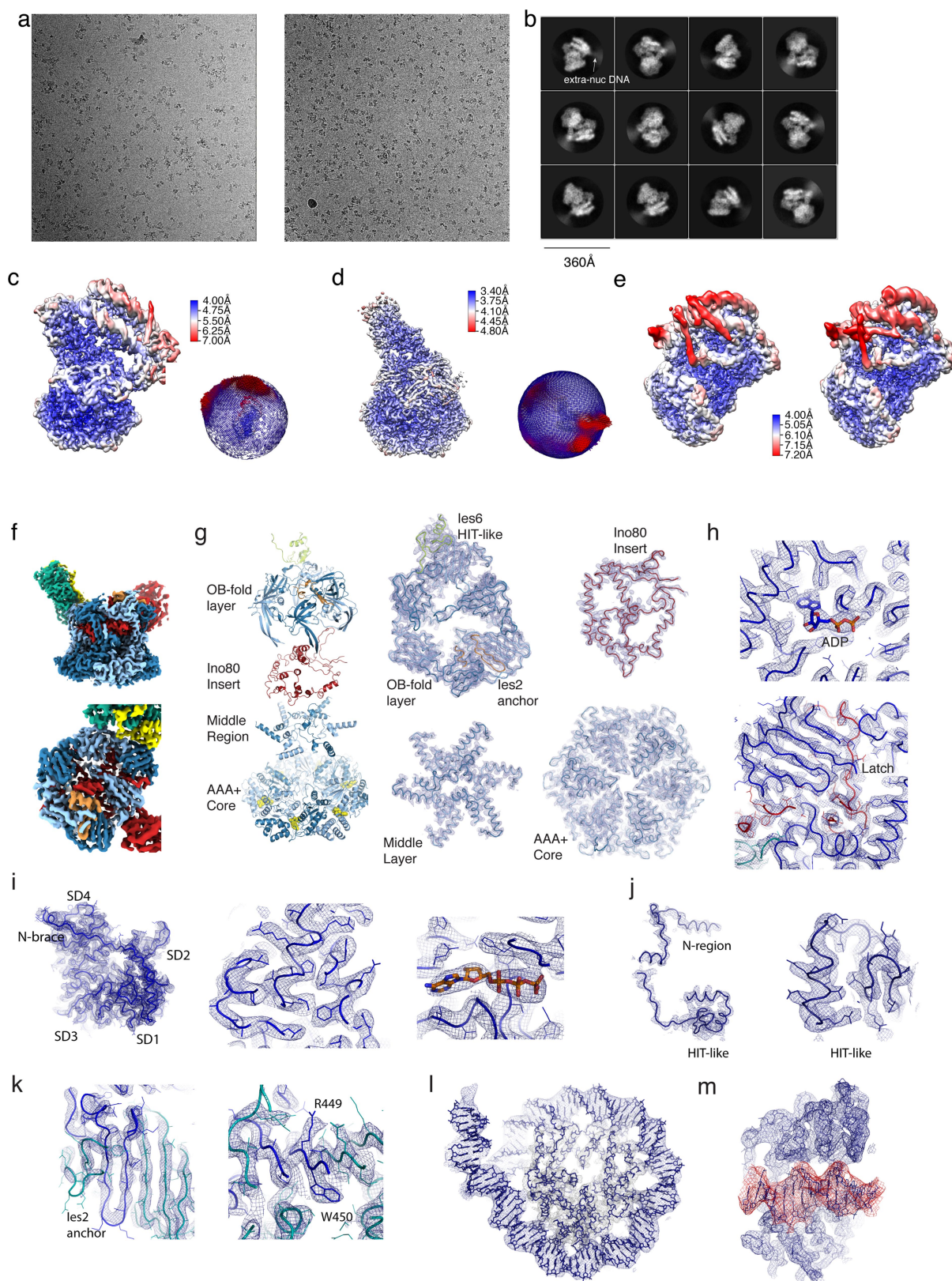
Extended Data Fig. 1 | Purification of apo INO80, INO80-ON50 and sliding activity of INO80. **a**, Schematic of expression and purification of INO80. **b**, SDS-PAGE of INO80 purification steps (stained with SimplyBlue). Protein identity was confirmed by mass spectrometry (data not shown). **c**, Quantification of band intensity from SDS-PAGE (SEC sample) plotted against the molecular weight shows stoichiometric presence of all subunits. **d**, Label-free semi-quantitative mass spectrometry analysis of INO80^{core} complexes after individual purification steps. **e**, Right-angle light scattering measurement of apo INO80. Measured refractive index and calculated logarithmical molecular weight are plotted

against the elution volume. The measurement yields a molecular weight of 880 kDa, confirming the integrity and correct stoichiometry of the purified complex. **f**, Comparison of the SEC elution profile of apo INO80 and the Arp5^{DBD} mutant on a Superose 6 3.2/300. **g**, Purification of the INO80-nucleosome complex. SEC elution profile from a Superose 6 3.2/300 is shown together with an analysis of the main peak fraction by SDS-PAGE. **h**, Sliding of end-positioned ON80 mononucleosomes by INO80. Native PAGE analysis of fluorescein-labelled nucleosome is shown. **i**, Interaction of INO80 and mononucleosome monitored by electrophoretic-mobility shift assay.



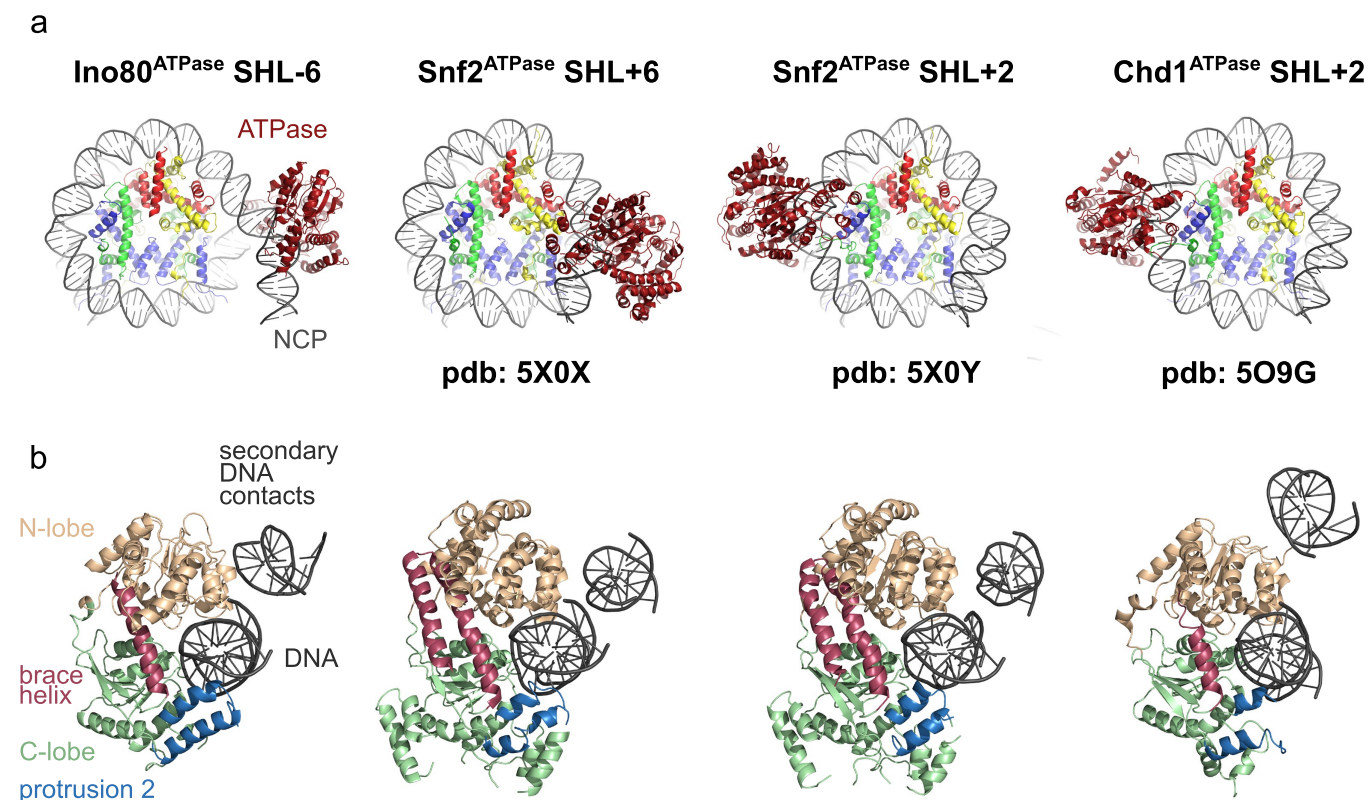
Extended Data Fig. 2 | Cryo-EM data analysis. a, b, Schemes of RELION⁴⁰ classifications and refinements that were used to obtain cryo-EM reconstructions of the INO80^{core}-NCP complex. **a,** Outline of an initial classification scheme that used a cryoSPARC⁴² ab initio 3D reconstruction of the complex as a reference. **b,** Classification scheme that yielded the final cryo-EM reconstructions. In **a** and **b**, boxed 3D classes were selected for further processing as indicated. Two-dimensional classes discarded for further processing are marked with an asterisk. **c,** Ab initio 3D reconstruction by cryoSPARC⁴² using 6,000 semi-automatically picked particles **d,** Eight hundred manually picked particles were used to obtain initial 2D classes that were used as references for automated particle picking as indicated in **a**. **e,** Projections of the experimentally

determined 5.8 Å cryo-EM reconstructions obtained from the scheme in **a**. These projections were low-pass filtered to 35 Å and used then as templates to improve automated picking of particles corresponding to sparsely populated orientations of the complex (see Methods). The quality of the automated particle picking was verified by visual inspection of micrographs as well as by diagnostic 2D classifications (not shown). Later 3D classifications in the scheme shown in **b** were facilitated by masks and fixed Euler angles from previous refinements as indicated (3D classification B3). **f,** Gold standard Fourier shell correlation curves of final maps (3.75, 4.34, 4.62 and 4.68 Å). The resolutions were determined using the 0.143 Fourier shell correlation criterion as indicated by the dotted line. Extended Data Table 1 summarizes data collection and processing.



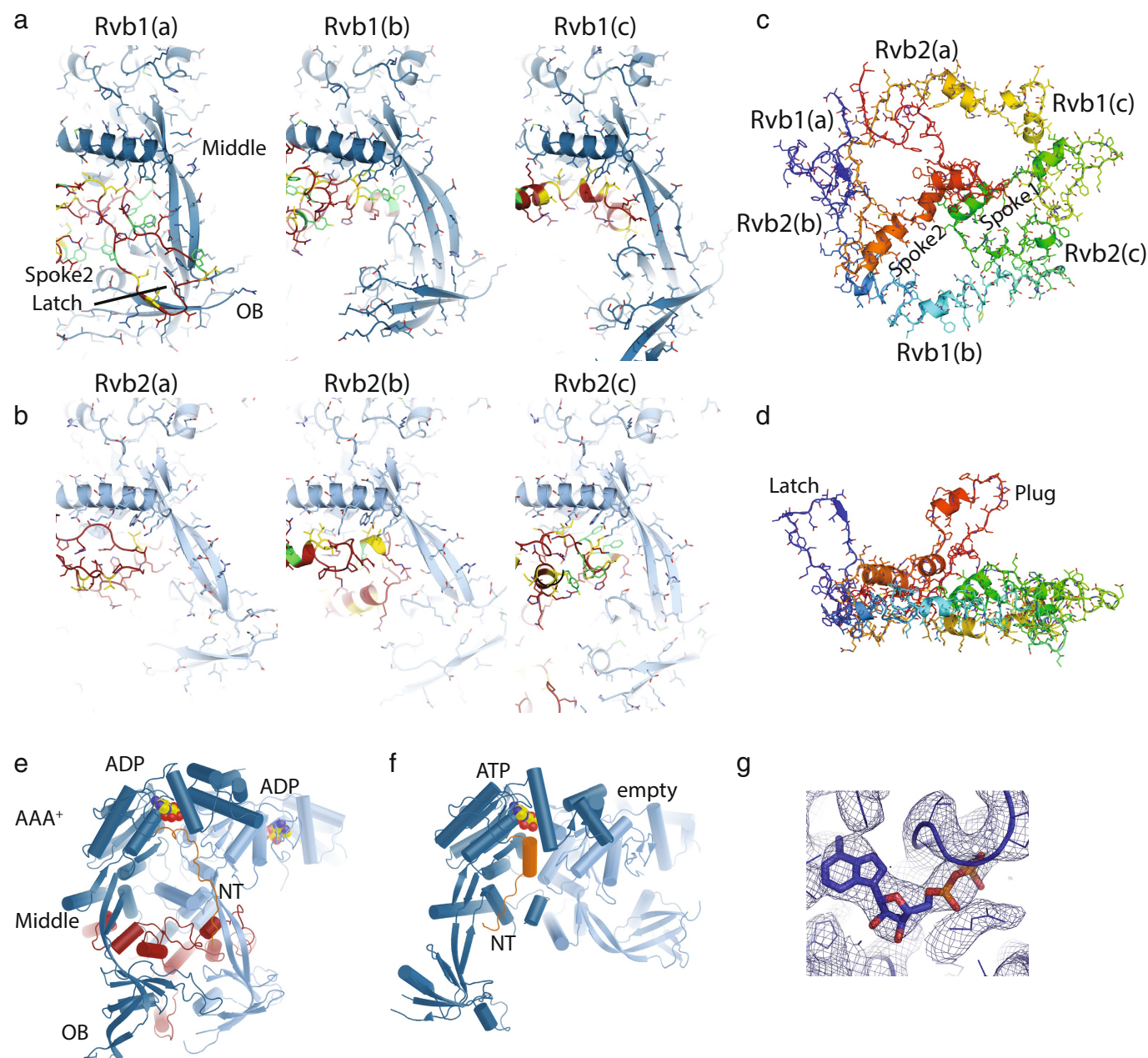
Extended Data Fig. 3 | Cryo-EM data quality. **a**, Two representative micrographs of the set that was used to determine the structure of the INO80^{core}-NCP complex. **b**, Typical 2D class averages of the INO80^{core}-NCP complex. Note that dynamic extranucleosomal DNA (extra-nuc DNA) visibly protrudes from the well-ordered core complex. **c–e**, The final 4.3 Å (**c**, overall), 3.7 Å (**d**, Rvb1/Rvb2–Arp5 mask) and 4.6 Å and 4.7 Å (**e**, grappler conformations A (right) and B (left)) maps were analysed by using ResMap⁵⁰. Local resolution estimates are shown as a colour-coded surface representation along with representations of angular distributions of particles contributing to the 4.3 and 3.7 Å maps. **f–m**, Representative examples of cryo-EM map areas used for model building. **f**, The 3.7 Å

map using the colour codes of Fig. 1c showing the definition of Rvb1/Rvb2–client interactions. **g**, ‘Explosion’ figure of the Rvb1/Rvb2 layers, along with corresponding regions of the 3.7 Å map. **h**, Top, details showing a representative ATP/ADP-binding site of Rvb1/Rvb2 with highlighted ADP, and showing the latch of the Ino80^{insert} (red). **i**, Map area at the Arp5^{core} showing the N-terminal brace (left), with representative details of the actin core (middle) and the ATP-binding site (right). **j**, Overview showing Ies6 (left) and details of its HIT-like domain (right). **k**, Map area at the Ies2–Rvb1/Rvb2 interaction (left) with details showing an anchoring tryptophane. **l**, Map area at the NCP. **m**, Map area at the Ino80 motor domain bound to SHL –6 (red).



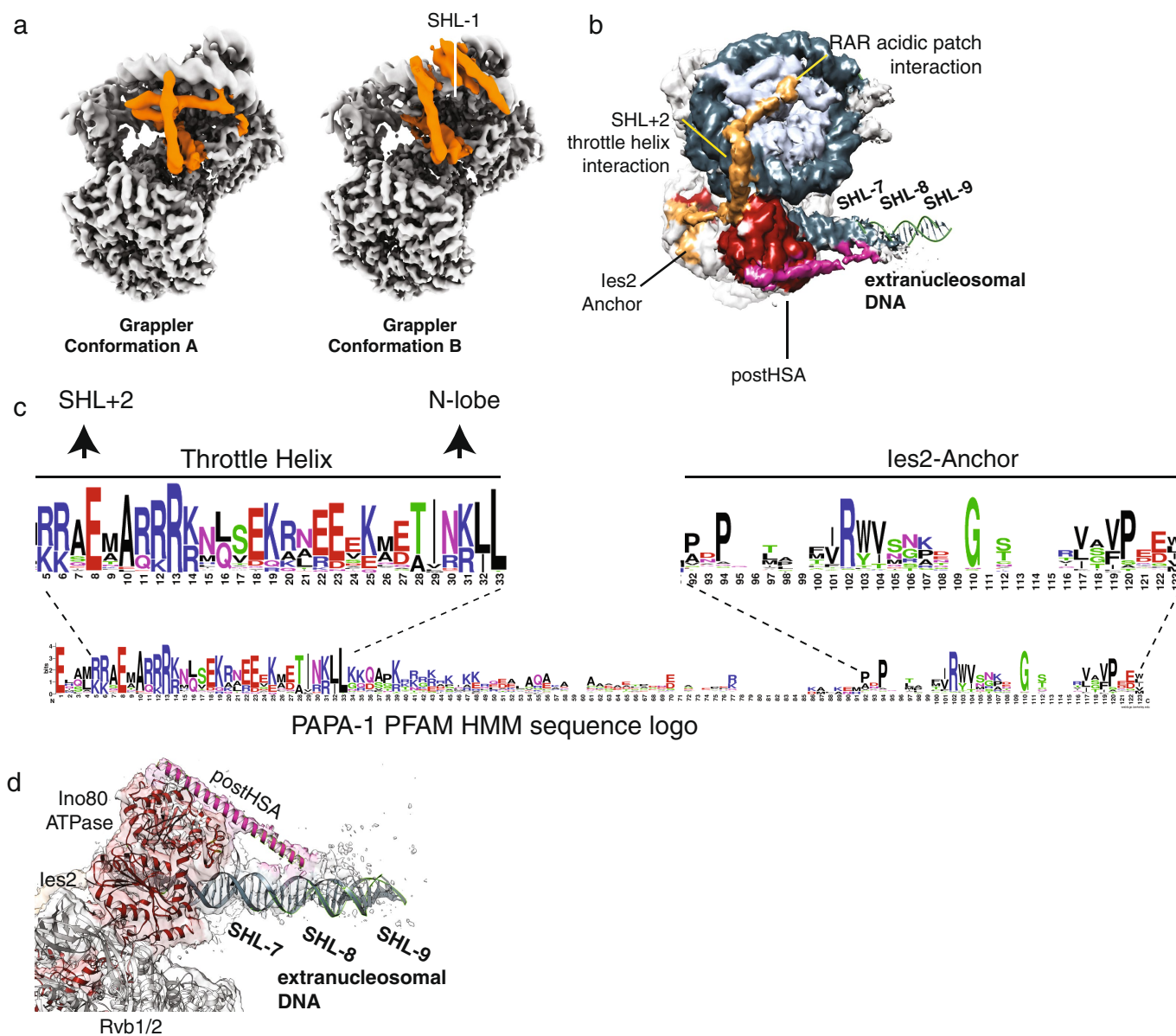
Extended Data Fig. 4 | Comparison of nucleosome-bound Swi2/Snf2-type ATPases. a, Interaction of Ino80^{ATPase} bound to SHL -6 (left, this study), Snf2^{ATPase} bound to SHL + 6 (middle), Snf2^{ATPase} bound to SHL + 2

and Chd1^{ATPase} bound to SHL + 2 (right) with NCPs. **b**, Comparison of domain architectures of the Swi2/Snf2-type ATPases and their interaction with nucleosomal DNA.



Extended Data Fig. 5 | Details of Rvb1/Rvb2–Ino80^{insert} interactions. **a**, Close-up views of Rvb1 client cavities (blue), bound to the different interaction elements of Ino80^{insert} (red, with yellow hydrophobic and green aromatic side chains). **b**, As in **a** but depicting Rvb2 client cavities. **c**, Ino80^{insert} shown in rainbow colouring from N terminus (red) to C terminus (blue), to highlight the circular fold. Selected elements as well as the positions of the Rvb1/Rvb2 binding partners are annotated. **d**, As in **c** but viewed from the side to highlight the protruding plug and latch elements. **e**, **f**, Rvb1/Rvb2 pair (the pair 1c and 2c from the hexamer in

Fig. 1c) bound to Ino80^{insert} (**e**) compared with a Rvb1/Rvb2 pair from the unliganded dodecameric state (**f**) (PDB code: 4WVY). The comparison shows how client binding arranges the AAA⁺, OB and middle layers and displaces the N-terminal domain of Rvb1 from the client pocket, also seen for human INO80^{core}²³. Both types of conformational changes have an effect on the ADP-binding site (ADP and ATP represented by colour-coded spheres), which suggests how client interactions are allosterically coupled to the ATPase activity of Rvb1/Rvb2. **g**, Exemplary view of the ADP coordination along with the superimposed map.



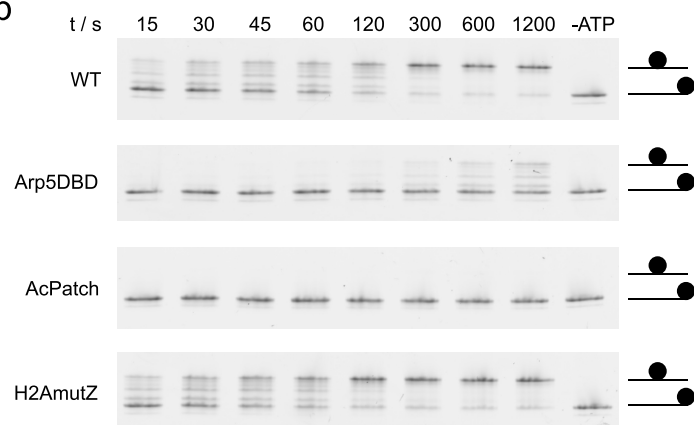
Extended Data Fig. 6 | Two conformations of the grappler element and location of the post-HSA domain. Masked 3D classifications identified two conformations of the grappler element of $\text{INO80}^{\text{core}}$ and the post-HSA domain of the $\text{Ino80}^{\text{ATPase}}$. **a**, Left, grappler conformation A (conformation discussed in this study). Right, open conformation B in which the bar interacts with SHL -1 of the nucleosome. **b**, Subclass showing the post-

HSA domain (magenta) at the $\text{Ino80}^{\text{ATPase}}$ (red). Post-HSA domain protrudes towards extranucleosomal DNA. Ies2 is depicted in orange. **c**, Hidden Markov model (HMM) sequence logo of Ies2, showing high sequence conservation at key Ino80 and Rvb1/Rvb2 interaction sites. **d**, Detailed view of the map around post-HSA domain and extranucleosomal DNA, with superimposed models.

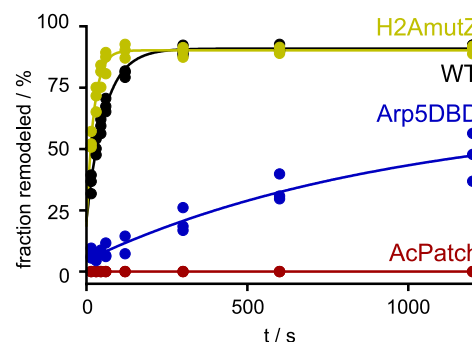
a

| | | | | | | |
|---------------------------------------|----|--------|----|---------|-------|-----|
| <i>Homo sapiens</i> H2A | 72 | DNKKTR | I | PRHLQLA | IRNDE | 91 |
| <i>Homo sapiens</i> H2A.Z | 75 | DLKVKR | IT | PRHLQLA | IRGDE | 94 |
| <i>Chaetomium thermophilum</i> H2A | 73 | DNKKTR | I | PRHLQLA | IRNDE | 92 |
| <i>Chaetomium thermophilum</i> H2A.Z | 85 | DLKVKR | IT | PRHLQLA | IRGDE | 104 |
| <i>Saccharomyces cerevisiae</i> H2A | 73 | DNKKTR | I | PRHLQLA | IRNDD | 92 |
| <i>Saccharomyces cerevisiae</i> H2A.Z | 80 | DLKVKR | IT | PRHLQLA | IRGDD | 99 |

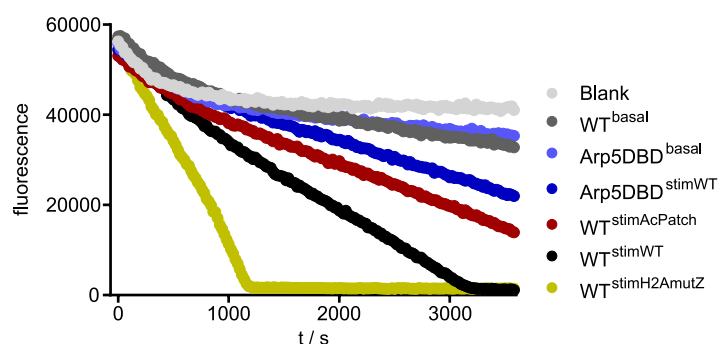
b



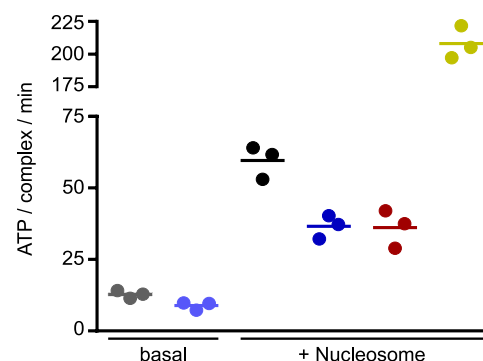
c



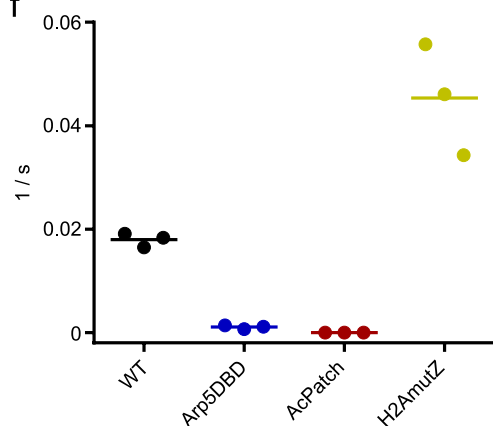
d



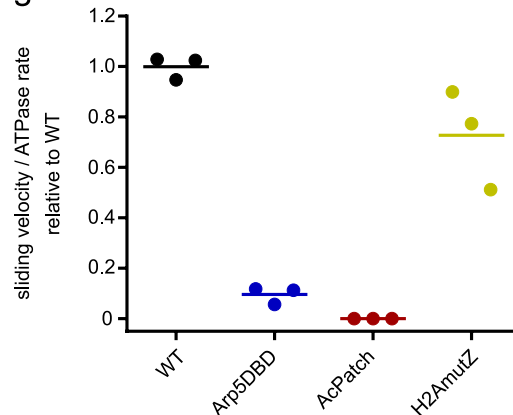
e



f



g



Extended Data Fig. 7 | Analysis of the enzymatic activity of INO80.

a, Sequence alignment of H2A and H2A.Z. Olive, residues at the interface of H2A with the foot of the grapples differ in a species-conserved fashion from H2A.Z. **b**, Sliding of 0N80 mononucleosomes by INO80 analysed by native PAGE. In the Arp5^{DBD} mutant, K88, R90, R92, K96, R112 and R118 are mutated to alanines. AcPatch (E61A, E64A, D72A and D90A) and H2AmuZ (N73L and N89G) describe mutants of grapples-contacting residues of H2A (see Fig. 3). Individual data points with exponential fit ($n=3$, technical replicates). **c**, Evaluation of the sliding activity of INO80. Band intensities of remodelled and unremodelled nucleosome species were quantified and the fraction of remodelled nucleosome plotted against time. Data points were fitted using an exponential equation. **d**, Raw data of ATPase assays. Basal

ATPase rates were determined for INO80 wild type (WT) and the Arp5^{DBD} mutant, along with nucleosome-stimulated rates. Superscripted text indicates whether a nucleosome was used to stimulate ATPase activity, and if so what type of nucleosome was used. **e**, ATPase rates of INO80 with and without stimulation by nucleosomes. Rates were calculated from the linear area of the raw data and were corrected for a buffer blank (colour code as in **d**). Mean and individual data points ($n=3$, technical replicates). **f**, Initial sliding rates of INO80 and mutants (colour code as in **c**). Data were derived from exponential fits of individual sliding curves in **c**. Mean and individual data points ($n=3$, technical replicates). **g**, Quotient of the sliding rate in **f** and ATPase rate in **e** normalized to the wild type. Mean and individual data points ($n=3$, technical replicates).

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

| | #1 INO80core (EMD-4264) (PDB 6FHS) | #2 INO80core+nuc (EMD-4277) (PDB 6FML) | #3 INO80 Grappler A (EMD-4278) | #4 INO80 Grappler B (EMD-4280) |
|---|--|--|-----------------------------------|-----------------------------------|
| Data collection and processing | | | | |
| Camera | Gatan K2 | Gatan K2 | Gatan K2 | Gatan K2 |
| Voltage (kV) | 300 | 300 | 300 | 300 |
| Electron exposure (e-/Å ²) | 59.6 | 59.6 | 59.6 | 59.6 |
| Defocus range (μm) | 1.3-3.5 | 1.3-3.5 | 1.3-3.5 | 1.3-3.5 |
| Pixel size (Å) | 1.06 | 1.06 | 1.06 | 1.06 |
| Symmetry imposed | C1 | C1 | C1 | C1 |
| Initial particle images (no.) | 251692 | 251692 | 251692 | 251692 |
| Final particle images (no.) | 144278 | 33937 | 16287 | 17263 |
| Map resolution (Å) | 3.75 | 4.34 | 4.68 | 4.62 |
| 0.143 FSC threshold | | | | |
| Map sharpening B-factor (Å ²) | -142 | -114 | -124 | -136 |
| Refinement | | | | |
| Initial model used (PDB code) | models for refinement were built <i>de novo</i> or based on 4WW4 (Rvb1/2) and 1NWK (actin) | models for refinement were based on 6FHS (INO80core) and 4R8P (nucleosome) | | |
| Model resolution (Å) 0.5 FSC threshold | 3.81 | 4.47 | | |
| Model resolution range (Å) | 360.4 - 3.7 | 360.4 - 4.3 | | |
| Map sharpening B factor (Å ²) | -142 | -114 | | |
| Model composition | | | | |
| Non-hydrogen atoms | 27383 | 43462 | | |
| Protein and DNA residues | 3510 | 5114 | | |
| Ligands (ADP and ATP) | 7 | 7 | | |
| B factors (Å ²) | | | | |
| Protein and DNA | 62.6 | 112.9 | | |
| Ligand | 57.8 | 80.3 | | |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.005 | 0.010 | | |
| Bond angles (°) | 0.82 | 1.20 | | |
| Validation | | | | |
| MolProbity score | 1.6 | 1.7 | | |
| Clashscore | 4.3 | 4.8 | | |
| Poor rotamers (%) | 0.0 | 0.1 | | |
| Ramachandran plot | | | | |
| Favored (%) | 93.4 | 93.3 | | |
| Allowed (%) | 6.6 | 6.6 | | |
| Disallowed (%) | 0.0 | 0.1 | | |

Structure and regulation of the human INO80–nucleosome complex

Rafael Ayala^{1,2}, Oliver Willhoft^{1,2}, Ricardo J. Aramayo¹, Martin Wilkinson¹, Elizabeth A. McCormack¹, Lorraine Ocloo¹, Dale B. Wigley^{1*} & Xiaodong Zhang^{1*}

Access to DNA within nucleosomes is required for a variety of processes in cells including transcription, replication and repair. Consequently, cells encode multiple systems that remodel nucleosomes. These complexes can be simple, involving one or a few protein subunits, or more complicated multi-subunit machines¹. Biochemical studies^{2–4} have placed the motor domains of several chromatin remodellers in the superhelical location 2 region of the nucleosome. Structural studies of yeast Chd1 and Snf2—a subunit in the complex with the capacity to remodel the structure of chromatin (RSC)—in complex with nucleosomes^{5–7} have provided insights into the basic mechanism of nucleosome sliding performed by these complexes. However, how larger, multi-subunit remodelling complexes such as INO80 interact with nucleosomes and how remodellers carry out functions such as nucleosome sliding⁸, histone exchange⁹ and nucleosome spacing^{10–12} remain poorly understood. Although some remodellers work as monomers¹³, others work as highly cooperative dimers^{11, 14, 15}. Here we present the structure of the human INO80 chromatin remodeller with a bound nucleosome, which reveals that INO80 interacts with nucleosomes in a previously undescribed manner: the motor domains are located on the DNA at the entry point to the nucleosome, rather than at superhelical location 2. The ARP5–IES6 module of INO80 makes additional contacts on the opposite side of the nucleosome. This arrangement enables the histone H3 tails of the nucleosome to have a role in the regulation of the activities of the INO80 motor domain—unlike in other characterized remodellers, for which H4 tails have been shown to regulate the motor domains.

We prepared a complex between human INO80 core complex¹⁰ and human nucleosomes flanked by 52 and 25 base pair overhangs (Extended Data Fig. 1) in the presence of ADP•BeF₃, which tightens nucleosome binding (Extended Data Fig. 1). Although this complex was prepared at a INO80:nucleosome molar ratio of 2:1, the majority of the particles on our electron microscopy grids contained either free INO80 complex or a 1:1 complex (Extended Data Fig. 2 and Methods). We processed the data to obtain two different reconstructions (Extended Data Fig. 2 and Methods). One was selected to obtain nucleosome complexes (4.8 Å resolution; Fig. 1, Extended Data Figs. 2, 3, Extended Data Table 1 and Supplementary Video 1). The other (3.8 Å resolution; Extended Data Figs. 2, 3 and Extended Data Table 1) initially used all particles, but during the final stages of processing the region corresponding to the bound nucleosome was masked out to optimize fitting on the INO80 component (Extended Data Fig. 2 and Methods). This map showed essentially the same features as our previous apo structure¹⁶ but with considerable improvement in areas such as IES2 (encoded by *INO80B*) and the RUVBL1–RUVBL2 hexamer¹⁶, enabling us to improve our model and assign sequence to the INO80-I region (Extended Data Fig. 4 and Methods). Furthermore, this map enabled us to determine the location of a zinc-binding domain of IES2. Parts of IES2 track across the RUVBL1–RUVBL2 hexamer and interact with the oligonucleotide/oligosaccharide-binding domains from

adjacent RUVBL1 and RUVBL2 subunits (Extended Data Fig. 4). This part of human IES2 corresponds to an extension at the C terminus that is absent in the yeast protein. The density runs towards the motor domains but is disordered beyond the interface with the RUVBL subunits. Previous crosslinking data from the yeast apo Ino80 complex indicated an interface between the Ies2 subunit and the motor domains¹⁷, and the IES2 subunit regulates ATPase activity in both yeast and human INO80 complexes^{10, 18–20}. The crosslinks observed between yeast Ies2 and the Ino80 motor domains are, in the human IES2 structure, located just beyond the ordered part of the structure but close to the motor domains (Extended Data Fig. 4).

The structure of the INO80–nucleosome complex reveals protein secondary structural elements (Extended Data Fig. 4) and a bound nucleosome (Fig. 1a–c). The RUVBL1–RUVBL2 heterohexamer

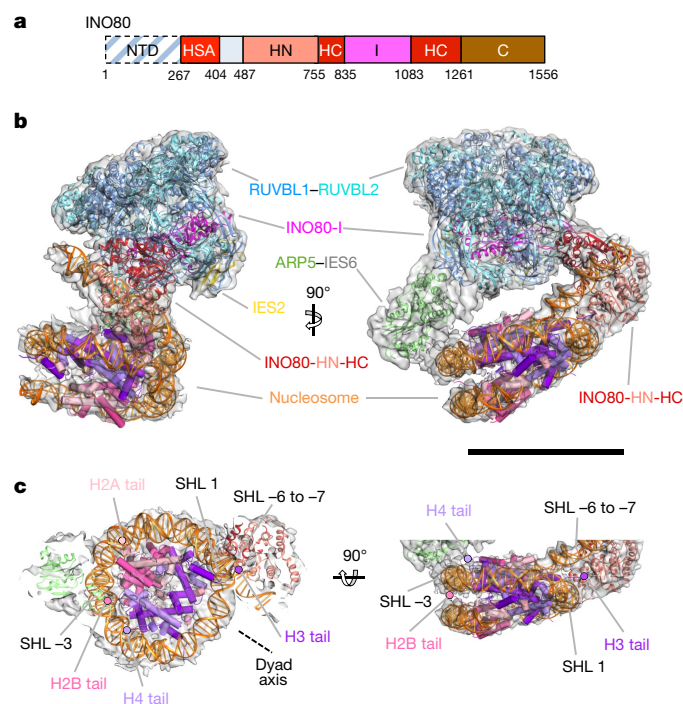


Fig. 1 | Human INO80–nucleosome complex. **a**, INO80 subunit with functional domains labelled. **b**, Three-dimensional INO80–nucleosome complex reconstruction with RUVBL1–RUVBL2, INO80, ARP5, IES2 and nucleosome structural models fitted. Scale bar, 100 Å. **c**, INO80–nucleosome interactions with histones and nucleosome positions labelled. INO80 contacts the nucleosome at SHL –6 and SHL –3. The locations of the histone tails are also shown. NTD, N-terminal domain; HSA, helicase–SANT-associated domain; HN, N-terminal helicase domain; HC, C-terminal helicase domain; I, INO80 insert domain; C, C-terminal domain.

¹Section of Structural Biology, Dept. Medicine, Imperial College London, London, UK. ²These authors contributed equally: Rafael Ayala, Oliver Willhoft. *e-mail: d.wigley@imperial.ac.uk; xiaodong.zhang@imperial.ac.uk

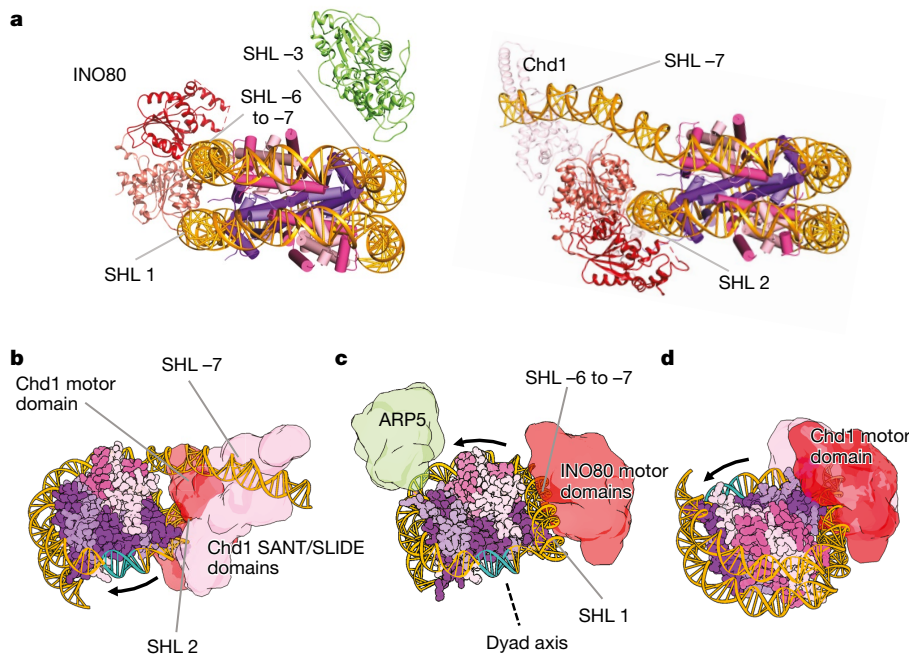


Fig. 2 | Comparison of INO80 with Chd1 and a model for translocation by INO80. **a**, INO80–nucleosome and Chd1–nucleosome complexes viewed from the side of the nucleosome. **b**, Chd1 is proposed to push DNA towards the dyad axis (cyan). **c**, INO80 (aligned on the nucleosome

as in **b**) would push DNA past ARP5–IES6 towards the dyad, but from the opposite direction. **d**, As in **b**, but with the view aligned on the motor domains as in **c**.

encloses a large insertion in the C-terminal INO80 motor domain (Fig. 1a, b and Extended Data Fig. 4), as also seen in a previously published INO80 apo structure¹⁶. In the nucleosome-bound complex, this insertion region is connected to a region of density that is much better ordered than in the apo complex and fits the C-terminal motor domain along-side (Extended Data Fig. 4). Consistent with binding of ADP•BeF₃, we observe the ATPase domains in the closed, nucleotide-bound state⁵. However, rather than being located at superhelical location (SHL) 2 of the nucleosome wrap, as previously observed in Chd1 and Snf2^{5,6}, the motor domains—as predicted from biochemical studies³—are instead located across SHL –6 and SHL –7 in an orientation consistent with tracking along one strand of the DNA duplex in the anticipated 3′–5′ direction (Fig. 2a and Extended Data Fig. 5), when compared to other well-characterized superfamily 2 DNA translocases such as NS3²¹. This orientation would pump duplex DNA from the overhang onto the nucleosome towards the dyad axis. This contact region for the motor domains differs completely from that of all other characterized remodellers (Fig. 2a and Extended Data Fig. 5), but is consistent with footprinting and crosslinking studies of the yeast Ino80 complex³. The INO80 footprint that spans SHL –6 and –7 is due to contacts with the motor domains. The N-terminal motor domain also contacts across the gyres at SHL 1. Similar contacts across the gyres are observed in the Snf2 and Chd1 structures^{5,6} and are essential for nucleosome sliding^{4,6}.

Notably, footprinting studies also indicated contacts at SHL –2 and –3³. Our structure reveals these to be due to ARP5–IES6 (encoded by *ACTR5* and *INO80C*, respectively) and are proximal to histones H2A and H2B, on almost the opposite side of the nucleosome to the contacts made by the motor domains (Fig. 1b, c, 2a). Consistent with these contacts, ARP5 binds to H2A–H2B dimers in solution and the ARP5–IES6 complex binds to nucleosomes (Extended Data Fig. 6). The ARP5–IES6 module also has a key role in coupling ATPase and sliding activities^{10,19,20}.

Although our structures reveal much detail about how INO80 contacts its nucleosome substrate, an obvious omission from them is the N-terminal region of the INO80 complex that contains the actin, ARP4 (encoded by *ACTL6A*) and ARP8 (encoded by *ACTR8*) subunits.

Notably, this region—termed SC1—is flexible in the apo structure, but careful selection of particles enabled us to locate this region of the complex¹⁶. This region remains flexible in the complex with nucleosomes, which results in it being averaged out in the structure. However, although it is visible in single particles and in carefully selected 2D class averages (Extended Data Fig. 7), it is too variable in location to be defined. This suggests that it does not make extensive contacts with the nucleosome in this conformational state. The SC1 components have previously been shown to interact with histones^{22,23} and it may be that this component also interacts with the histone core in the active INO80 dimer or in a different functional state on the catalytic pathway.

INO80, Chd1 and Snf2-like enzymes all translocate duplex DNA by tracking principally along one strand with a 3′–5′ directionality^{8,24,25} in a manner analogous to that used by single-strand superfamily 2 translocases such as NS3²¹. However, whereas Ino80 and Chd1 slide nucleosomes away from DNA ends^{10,12}, Snf2-like enzymes instead slide nucleosomes towards DNA ends²⁵. Although similar regions of nucleosomal DNA are contacted in each case, our structures place the motor domains of INO80 at a different location than those in Chd1 and Snf2 (Fig. 2a and Extended Data Fig. 5). A consequence of this difference is that INO80 would pump DNA from the overhang towards the dyad, whereas Chd1 and Snf2 would do this from the opposite direction^{5,6} (Fig. 2 and Extended Data Fig. 5). This position of the motor domains of INO80 would move nucleosomes away from ends, consistent with biochemical observations^{10,12}. The common directionality of sliding towards DNA ends suggested by the Chd1 and Snf2 structures raises a conundrum, because Chd1 and Snf2-like enzymes have previously been shown to have opposing directional specificities for nucleosome sliding^{24,25}.

The Snf2–nucleosome and Chd1–nucleosome complexes show broadly similar contacts between their motor domains and the SHL 2 position of the DNA wrap. Both structures also show contacts across the DNA gyres to contact SHL –6, as predicted by biochemical studies⁴. Previous work has shown that even the closely related Swr1 complex—which shares some subunits with the INO80 complex—is positioned at SHL 2 to SHL 3²⁶. By contrast, the motor domains of INO80 bind at a completely different location but still contact the DNA across the gyres, albeit at different parts of the nucleosome wrap.

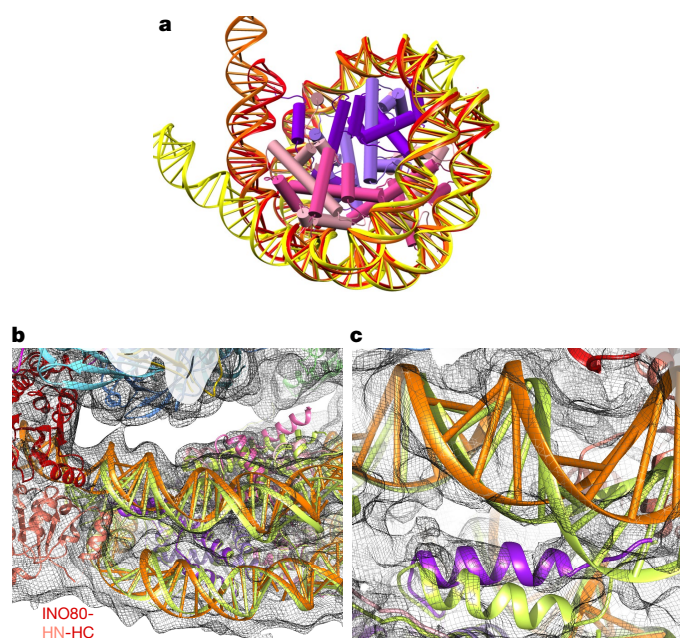


Fig. 3 | Nucleosome distortion in the INO80–nucleosome complex. **a**, DNA is peeled off in INO80 (orange) and Chd1 (yellow) compared to the free nucleosome (red). In INO80 this is due to motor domain interaction, whereas in Chd1 this is due to the SANT and SLIDE domain interactions. **b**, DNA near the motor domains in INO80–nucleosome (orange) is lifted compared to a canonical nucleosome (light green), which also causes a slight rotation of H2A–H2B (pink). **c**, Lifting of the DNA in INO80–nucleosome complex and movement of the H3 N-terminal helix (purple) near the motor domains.

The binding of INO80 induces unwrapping of the DNA at SHL –6 to SHL –7, though to a lesser extent than does Chd1 binding (Fig. 3a and Extended Data Fig. 5). However, the consequences of INO80 binding are more marked because of a more-subtle distortion of the DNA wrap that extends all the way from the motor domains to the ARP5–IES6 contact. The distortion lifts one DNA gyre away from the other. The associated H2A–H2B dimer moves along with the DNA, which causes the dimer to lift away from the H3–H4 tetramer and presumably weakens this interface (Fig. 3b). Finally, as a consequence of the peeling back of the DNA at the entry site, the histone H3 tail remains associated with the DNA and alters conformation compared to the H3–H4 core (Fig. 3c). These conformational changes may have a role in histone exchange.

The ARP5–IES6 subunits couple ATP hydrolysis to nucleosome sliding in INO80^{10, 19, 20}. Furthermore, cyclic partial unwrapping of the DNA around the H2A–H2B interface is required for the histone exchange activity previously reported for INO80³. The location of the DNA contacts we observe here suggests a simple mechanism for such a process. The directionality of translocation by the motor domains would push DNA towards the ARP5–IES6 contact region (Fig. 2c). Unless released, this would result in a partial unwrapping of the DNA wrap, bulging out between these contacts across the H2A–H2B interface to facilitate H2A–H2B dimer exchange. Even though our structure has not undergone catalytic ATP turnover, the distortions induced by binding to INO80 appear to prepare the nucleosome for dimer exchange. Further ATP-dependent translocation by the motor domains would increase this effect by pushing DNA towards ARP5–IES6. The Swr1 complex, which is related to INO80 and contains several subunits in common with it, facilitates histone exchange but is unable to slide nucleosomes²⁷. Unlike those of INO80, the motor domains of Swr1 are located at the canonical SHL 2 position and the Swc2 subunit contacts the DNA overhang²⁸. This two-point contact, with motor domain and DNA overhang contacts swapped relative to the INO80 complex, raises the possibility that a similar mechanism releases the H2A–H2B dimer

from its DNA interface, with the motor domains pushing (or pulling) against a second contact to provide strain and lift the DNA wrap from the nucleosome surface.

INO80 slides nucleosomes from DNA ends and is able to sense a flanking DNA length of up to 50–60 bp^{12, 14}. As with some other remodellers^{11, 15}, INO80 acts as a cooperative dimer in sliding¹⁴. ATPase activity becomes uncoupled from sliding when INO80 has positioned nucleosomes at the centre of short DNA fragments, but continues at the same rate as when the nucleosome is sliding^{12, 14}. The two contact points with the nucleosome in the structure suggest a basis for this behaviour. Because the motor domains pump DNA towards the dyad via the ARP5–IES6 contact, they will be underwinding the DNA as well as unwrapping it from the nucleosome surface. If the motor domains were to slip, which could happen when the DNA overhang becomes too short, then the DNA could simply re-associate with the nucleosome surface, which would result in a futile cycle of ATP hydrolysis. On the other hand, if the grip by ARP5–IES6 were to slip the DNA could be pushed forward across the surface, resulting in sliding of the DNA wrap across the nucleosome surface. As a result, a translocation step size of one base per ATP—as shown for most SF1 and SF2 helicases²⁹—might build up tension in the DNA, before being released in apparently larger step sizes as DNA slips past the ARP5–IES6 grip point. Precisely such behaviour has previously been observed for nucleosome sliding at the single molecule level^{30, 31} and may be an intrinsic part of nucleosome remodelling mechanisms. Such a mechanism would be coupled to sliding were it to prevent ‘back-slippage’, and thus provide directional translocation against a ratchet. For enzyme systems that require dimers, a mechanism that regulates the forward slippage between the partners could explain this behaviour, which presumably correlates with some form of regulation of activity—particularly for remodellers such as INO80 that have higher-order functions such as nucleosome spacing and phasing^{12, 14, 32}.

Several remodelling complexes are regulated by histone H4 tails^{33–35} through a complex interplay between regulatory components (termed AutoN and NegC) of the motor domains that are missing in INO80¹⁴. The unique binding mode of INO80 raises questions about its regulation by histones because the H4 tails are too far away to interact with the motor domains of INO80 (Fig. 1c), which suggests regulation by a different mechanism. We prepared a number of nucleosome variants in which the histone tails were individually deleted (Fig. 4 and Extended Data Fig. 8). For these tailless nucleosome substrates, sliding rates were comparable in all variants tested. Because INO80 functions as a dimer¹⁴, we also assessed the effects of the histone tails on cooperativity. For the H3 tail deletion, the Hill coefficient dropped considerably both for activity and binding (Fig. 4a), demonstrating a contribution of the H3 tail to INO80 dimer cooperativity that we localized to residues 31–39 (Fig. 4b). The ATPase activity and affinity of INO80 for tailless and cognate nucleosomes was similar (Fig. 4c, d), consistent with its regulation being distinct from other remodellers that are regulated by H4 tails³⁴. Individual mutations to mimic lysine acetylation (K36Q and K37Q) both showed a small but reproducible stimulation of sliding activity, but the K37Q mutation also showed the loss in cooperativity observed with the full H3 tail truncation (Fig. 4e and Extended Data Fig. 9). A double mutation showed a cooperative effect in sliding while retaining the loss in cooperativity. By contrast, a control substitution (K27Q) showed no effect on activity. These data support a role for H3 tails in regulating cooperativity in INO80 sliding and identify K37 as a key component in this process. The location of one H3 tail adjacent to the motor domains supports this idea (Fig. 1c) but, rather than being adjacent to the C-terminal motor domain as seen for remodellers regulated by H4 tails^{5, 36}, the H3 tail instead sits next to the N-terminal motor domain of the INO80 subunit. The location of this H3 tail is normally between the DNA gyres, as one end exits the nucleosome wrap³⁷. However, the unwrapping of DNA from the nucleosome surface that we observe in our structure breaks these contacts at the DNA entry site causing the H3 tail at that site to undergo a conformational change in response. Evidently, this unwrapping is required to initiate sliding

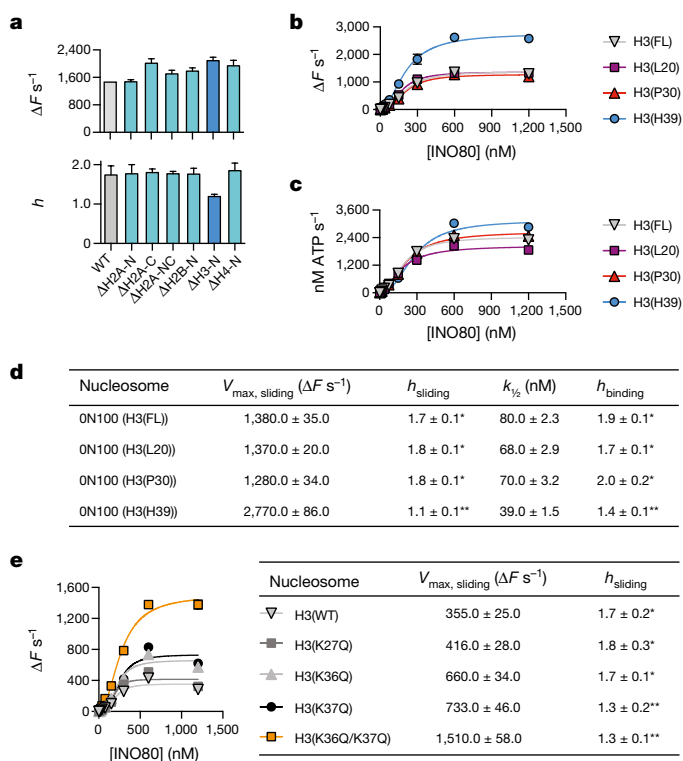


Fig. 4 | INO80 is regulated by H3 tails. **a**, Initial nucleosome sliding rates of human nucleosomes lacking different histone tails (-N, N-terminal tail; -C, C-terminal tail; -NC, N- and C-terminal tails), using a fluorescence resonance energy transfer-based assay²⁵. h refers to the Hill coefficient for cooperativity. **b**, Effect of increasing the extent of H3 tail truncation on nucleosome sliding. No effect is observed with 30 residues removed but a 39-residue truncation induces stimulation of sliding. H3(FL), full length; H3(L20), residues 1–20 removed; H3(P30), residues 1–30 removed; H3(H39), residues 1–39 removed. **c**, ATPase rates for H3 tail truncations. **d**, Loss of cooperativity for both sliding and binding of nucleosomes. ON100, nucleosome flanked by 0 and 100 base pair overhangs. $k_{1/2}$ refers to the half-saturation point for nucleosome binding. **e**, Lysine to glutamine mutations in the H3 tail affect both the rate and cooperativity of sliding. $n = 2$ biologically independent experiments in all graphs. Error bars represent s.d. from the mean values. In **d**, * denotes $h > 1.5$ and ** denotes $h < 1.5$.

by INO80, although the details of this process require determining the structure of an INO80 dimer bound to a nucleosome.

Our work reveals that INO80 adopts a unique mode of interaction with nucleosomes that permits—or possibly requires—regulation by a mechanism that also differs from other systems. However, further work is required to determine details of these interactions and how these relate to the requirement of INO80 dimers for sliding activity.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0021-6>.

Received: 27 October 2017; Accepted: 16 February 2018.

Published online: 11 April 2018

- Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
- Dechassa, M. L. et al. Architecture of the SWI/SNF-nucleosome complex. *Mol. Cell. Biol.* **28**, 6010–6021 (2008).
- Brahma, S. et al. INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat. Commun.* **8**, 15616 (2017).
- Nodelman, I. M. et al. Interdomain communication of the Chd1 chromatin remodeler across the DNA gyres of the nucleosome. *Mol. Cell* **65**, 447–459 (2017).

- Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome–Chd1 structure and implications for chromatin remodelling. *Nature* **550**, 539–542 (2017).
- Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2–nucleosome structure. *Nature* **544**, 440–445 (2017).
- Sundaramoorthy, R. et al. Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes. *eLife* **6**, e22510 (2017).
- Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541–544 (2000).
- Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200–213 (2011).
- Willhoft, O., Bythell-Douglas, R., McCormack, E. A. & Wigley, D. B. Synergy and antagonism in regulation of recombinant human INO80 chromatin remodeling complex. *Nucleic Acids Res.* **44**, 8179–8188 (2016).
- Leonard, J. D. & Narlikar, G. J. A nucleotide-driven switch regulates flanking DNA length sensing by a dimeric chromatin remodeler. *Mol. Cell* **57**, 850–859 (2015).
- Udagama, M., Sabri, A. & Bartholomew, B. The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol. Cell. Biol.* **31**, 662–673 (2011).
- Qiu, Y. et al. The Chd1 chromatin remodeler shifts nucleosomal DNA bidirectionally as a monomer. *Mol. Cell* **68**, 76–88e76 (2017).
- Willhoft, O. et al. Crosstalk within a functional INO80 complex dimer regulates nucleosome sliding. *eLife* **6**, e25782 (2017).
- Racki, L. R. et al. The chromatin remodeler ACF acts as a dimeric motor to space nucleosomes. *Nature* **462**, 1016–1021 (2009).
- Aramayo, R. J. et al. Cryo-EM structures of the human INO80 chromatin-remodeling complex. *Nat. Struct. Mol. Biol.* **25**, 37–44 (2018).
- Tosi, A. et al. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–1219 (2013).
- Chen, L., Conaway, R. C. & Conaway, J. W. Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proc. Natl Acad. Sci. USA* **110**, 20497–20502 (2013).
- Watanabe, S. et al. Structural analyses of the chromatin remodelling enzymes INO80-C and SWR-C. *Nat. Commun.* **6**, 7108 (2015).
- Yao, W. et al. Assembly of the Arp5 (actin-related protein) subunit involved in distinct INO80 chromatin remodeling activities. *J. Biol. Chem.* **290**, 25700–25709 (2015).
- Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl Acad. Sci. USA* **107**, 521–528 (2010).
- Saravanan, M. et al. Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex. *Proc. Natl Acad. Sci. USA* **109**, 20883–20888 (2012).
- Gerhold, C. B. et al. Structure of actin-related protein 8 and its contribution to nucleosome binding. *Nucleic Acids Res.* **40**, 11036–11046 (2012).
- Stockdale, C., Flaus, A., Ferreira, H. & Owen-Hughes, T. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J. Biol. Chem.* **281**, 16279–16288 (2006).
- Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat. Struct. Mol. Biol.* **12**, 747–755 (2005).
- Ranjan, A. et al. H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. *eLife* **4**, e06845 (2015).
- Mizuguchi, G. et al. ATP-driven exchange of histone H2A.Z variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).
- Ranjan, A. et al. Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. *Cell* **154**, 1232–1245 (2013).
- Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
- Blosser, T. R., Yang, J. G., Stone, M. D., Narlikar, G. J. & Zhuang, X. Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* **462**, 1022–1027 (2009).
- Deindl, S. et al. ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* **152**, 442–452 (2013).
- Krietenstein, N. et al. Genomic nucleosome organization reconstituted with pure proteins. *Cell* **167**, 709–721.e712 (2016).
- Clapier, C. R., Längst, G., Corona, D. F., Becker, P. B. & Nightingale, K. P. Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol. Cell. Biol.* **21**, 875–883 (2001).
- Clapier, C. R. & Cairns, B. R. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* **492**, 280–284 (2012).
- Dang, W., Kagawa, M. N. & Bartholomew, B. Regulation of ISW2 by concerted action of histone H4 tail and extranucleosomal DNA. *Mol. Cell. Biol.* **26**, 7388–7396 (2006).
- Yan, L., Wang, L., Tian, Y., Xia, X. & Chen, Z. Structure and regulation of the chromatin remodeler ISWI. *Nature* **540**, 466–469 (2016).
- Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).

Acknowledgements We thank A. Siebert and Y. Chaban at eBIC for assistance with data collection, and C. Aylett for help in preparing the nucleosome model. Electron microscopy access and support were provided by the UK national

electron Bio-Imaging Centre (eBIC) (proposal EM14769), funded by the Wellcome Trust, MRC and BBSRC. The work was funded by the Wellcome Trust (D.B.W. and X.Z.), Cancer Research UK (D.B.W.) and an Imperial College President's PhD Scholarship (R.A.).

Reviewer information *Nature* thanks B. Bartholomew, O. Llorca and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.B.W. and X.Z. designed the studies. R.A., R.J.A., M.W. and O.W. performed the cryo-electron microscopy analysis, and built and refined the structural models. O.W., E.A.M. and L.O. prepared the samples. O.W. conducted the biochemical experiments. D.B.W. and X.Z. wrote the manuscript with input from all the authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0021-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0021-6>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.B.W. or X.Z.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Preparation of nucleosomes. For electron microscopy sample preparation, a 52N25 nucleosome was used (N refers to the 147-bp nucleosome core). Although we used the Widom-601 positioning sequence³⁸ as the basis for this core, we introduced a point mutation within the sequence to remove a *HinfI* restriction site (GATTC to GATTG) to assist with sample preparation. The nucleosome was then prepared via a previously described ligation method¹⁰. Tailless histones H2A-ΔN (A21–K130), H2A-ΔC (S1–L116), H2A-ΔNC (A21–L116), H2B-ΔN (K28–K125), H3-ΔN/H3(H39) (H39–A135), H3(P30) (P30–A135), H3(L20) (L20–A135), H4-ΔN (N25–G102), additional mutations in H3 (H3(K27Q), H3(K36Q), H3(K37Q) and H3(K36Q/K37Q)) and an H4(N25C) mutation for labelling were introduced by standard mutagenesis methods. Tailless-H3 and full-length nucleosomes were labelled on H4(N25C). Human H2A, H2B, H3.1 and H4 were co-expressed in *Escherichia coli*, lysed in buffer A (20 mM Tris pH 7.5, 400 mM NaCl, 0.1 mM EDTA, 1 mM TCEP) and purified as soluble octamers on HiTrap Heparin HP in buffer A and eluted with a salt gradient, followed by Superdex S200 in buffer B (20 mM Tris pH 7.5, 2 M NaCl, 0.1 mM EDTA, 1 mM TCEP). Following labelling with Alexa Fluor 555 or 647 C₂-maleimide, the octamer was re-purified by Superdex S200 in buffer B.

Preparation of INO80–nucleosome–ADP•BeF₃ complexes. INO80 complex was prepared as previously described¹⁰. Nucleosomes were prepared as described above. INO80–nucleosome–ADP•BeF₃ complexes were prepared at a final concentration of 350 nM INO80, 175 nM nucleosome, 3 mM ADP, 3 mM BeCl₂, 15 mM NaF and 5 mM MgCl₂. INO80, nucleosomes, ADP and MgCl₂ were prepared at 10× concentration in EM buffer (25 mM HEPES pH 8.0, 50 mM NaCl, 1 mM TCEP). BeCl₂ and NaF were prepared at 10× concentration in water. The components were then mixed in the following order. First, INO80 and nucleosomes were mixed together with the volume of EM buffer needed to obtain the final concentrations and incubated at 37°C for 15 min. This was followed by the addition of ADP and MgCl₂ and a further 15 min incubation at 37°C. Lastly, NaF and BeCl₂ were added simultaneously.

Electron microscopy grid preparation. Grids for cryo-electron microscopy were prepared by depositing 3.5 μl sample onto Quantifoil R2/2 copper grids. Samples were blotted before being flash-frozen in liquid ethane at liquid nitrogen temperature with a FEI Vitrobot Mark IV (waiting time 30 s, blotting time 0.5 s) at 4°C and 100% humidity.

Data collection. A set of 5,479 movies was collected at eBIC (Diamond Light Source) on a Titan Krios microscope operated at 300-kV acceleration voltage. Images were recorded on a Falcon 3EC direct electron detector operating in linear mode at a magnification of 129,000 for a final pixel size of 1.09 Å/pixel with defocus range from −2.0 to −4.0 μm. The total dose was 80 e[−]/Å² fractionated over 39 frames.

Image processing. Individual movie frames were aligned using MotionCor2³⁹. CTF parameters were estimated using Gctf⁴⁰. Particle picking was performed in Gautomatch using class averages obtained from a small dataset of the same sample previously collected in-house. Subsequent image processing was carried out in RELION 2.1.B1⁴¹ and cryoSPARC 0.5.6⁴². Global and local resolution estimates were calculated in RELION using the gold-standard Fourier shell correlation (FSC = 0.143) criterion⁴³. A total of 1,160,399 particles was extracted into boxes of 270 × 270 pixels. After 2D classification in cryoSPARC to remove false positives and noisy particles, a set of 775,804 particles was selected to perform downstream image processing, which is summarized in route A in Extended Data Fig. 1. In brief, nucleosome-bound particles were selected by a combination of 2D and 3D classifications in cryoSPARC and RELION. A final set of 26,416 homogeneous nucleosome-bound particles was selected to perform a final 3D refinement in RELION. The final model was refined to an overall resolution of 4.8 Å. FSC calculation was calculated after applying a mask generated by binarizing the map at a threshold of 0.012, extending the resulting mask by 6 pixels and adding a soft edge of 7 pixels. Statistics regarding the final model are presented in Extended Data Fig. 2 and Extended Data Table 1.

A 3.8 Å map was generated with 91,607 particles selected by 3D classification in RELION with a mask that excluded the nucleosome (route B in Extended Data Fig. 1). This map was used to aid model building.

Model building and refinement. Deposited coordinates for RUVBL1–RUVBL2 for the apo INO80 structure (RCSB Protein Data Bank (PDB) code: 5OAF) were docked into the nucleosome-bound map. These were then adjusted and manually rebuilt in COOT⁴⁴ with the aid of the 3.8 Å map. A homology model for ARP5 was generated by submitting the sequence to I-TASSER⁴⁵, using a series of actin-fold proteins as templates. Well-resolved secondary structure was built according to the density in the 4.8 Å INO80–nucleosome map. After adjustments the model was trimmed of all side chains. The sequence for IES2 was submitted to the PHYRE⁴⁶

server, which yielded multiple results for the C-terminal zinc-binding domain. A published structure of the zf-HIT domain of TRIP3 (PDB code: 2YQQ) was used as a starting model, and the coordinates were then manually extended towards the N terminus in COOT. A homology model for the INO80 motor domains was generated by threading the sequence into the structure of the Chd1 motor domain (PDB code: 5O9G) with SWISS-MODEL⁴⁷. Side chains were removed and the domains were rigid-body fit into the map, followed by a round of jellybody refinement in REFMAC⁴⁸. The INO80 insert domain was manually built in COOT and connected to the INO80 motor domains using the 3.8 Å map. Coordinates for a human nucleosome core particle (PDB code: 5AV9) were fit into the density corresponding to the nucleosome. Keeping the position of the histone octamer fixed, a model for the nucleosome was built in COOT by combining the coordinates of the human histone octamer and a Widom 601 DNA wrap (PDB code: 3LZ0). The region of DNA bound to the motor domains was extended using linear B-form DNA, following the path of DNA through Chd1 (PDB code: 5O9G) where possible. Histones H2A and H2B with their complexed DNA, as well as the N-terminal helix of H3, were moved according to clear changes in the density from the canonical position. After completing model building the coordinates were subject to real-space refinement in Phenix⁴⁹.

Purification of actin and actin-related proteins. Human actin and actin-related proteins (ARP5 and ARP8) were expressed in Hi5 insect cells with an N-terminal octahistidine and C-terminal double-Strep tag. All proteins were purified to near homogeneity using sequential affinity chromatography steps (HisTrap HP followed by StrepTactin HP (GE Healthcare)). This was followed by buffer exchange into a storage buffer containing 50 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP and 10% glycerol using a spin concentrator. The concentrated sample was then flash-frozen in liquid nitrogen in small aliquots until further use. ARP5–IES6 was prepared as previously described¹⁰.

Actin and actin-related protein pulldown assay. Purified human actin or actin-related proteins (bait) and recombinant human H2A–H2B dimers (prey) were prepared at concentrations of 20 and 40 μM, respectively, in pulldown buffer (25 mM Tris-HCl, 250 mM NaCl, 1 mM TCEP, 0.05% NP40). For each pulldown condition, these 2× stocks were mixed in equal volumes and placed on a roller at room temperature for 30 min to equilibrate. In preparation for the pulldown, 50 μl Strep-Tactin Magnetic Beads slurry (Qiagen) was washed with pulldown buffer on a magnetized support stand. After incubation, the protein mixture was added to the washed magnetic beads and incubated for a further 30 min at room temperature. The resin was then washed extensively with pulldown buffer (at least ten 1-ml washes) to remove any unbound products before adding 50 μl SDS-containing loading dye and boiling the sample. The bound products (that is, those eluted from the resin following addition of loading dye) were then resolved by SDS–PAGE.

ARP5 and ARP5–IES6 electrophoretic-mobility shift assay with nucleosomes. Purified ARP5 and ARP5–IES6 were incubated for 30 min at room temperature with 2 μM human 167 nucleosomes in buffer containing 25 mM Tris-HCl, 100 mM NaCl and 1 mM DTT. Final concentrations of ARP5 and ARP5–IES6 were as indicated in Extended Data Fig. 5. Equilibrated samples were then resolved by native PAGE on 6% acrylamide gels prepared and run in 0.5× TBE (Tris-borate-EDTA) buffer.

Microscale thermophoresis with ARP5–Ies6 and various substrates. Microscale thermophoresis (MST) experiments were carried out similarly to those previously described¹⁴, for the interaction of the INO80 core complex and nucleosomes. In brief, ARP5–IES6 was assayed for interaction with 0N100 nucleosomes, DNA (100 bp) and H2A–H2B histone dimers. ARP5–IES6 was incubated at the appropriate concentrations with 40 nM fluorescently labelled substrate for 30 min at room temperature in buffer containing 25 mM HEPES pH 8.0, 50 mM NaCl, 1 mM TCEP, 10% glycerol, 0.1 mg/ml BSA and 0.01% Tween-20. Reactions were loaded into Premium Coated Capillaries (Nanotemper) and analysed using a Monolith NT.115 (Nanotemper). Thermophoresis data were extracted from the companion software and analysed in Prism 6 (Graphpad) graphing software with a 'One site – specific binding with Hill slope' model. Nucleosomes were labelled on H4(N25C) as previously described¹⁰.

Nucleosome sliding assays. Increasing concentrations of INO80 were incubated with 6 or 18 pmol end-positioned nucleosomes with 100 bp flanking DNA for 15 min at 37°C in a 54 μl volume in buffer containing 25 mM HEPES pH 8.0, 50 mM NaCl and 1 mM TCEP. Following incubation, 45 μl of these reaction mixes were transferred into a 384-well microtitre plate. Reactions were initiated by injection of 5 μl ATP and MgCl₂ to a final concentration of 1 and 2 mM, respectively. Initial rate comparisons between full length and tailless nucleosomes were made by monitoring a change in FRET between Alexa Fluor 647 (Thermo Fisher Scientific) on the short-end of the DNA wrap, and Alexa Fluor 555 C₂-maleimide (Thermo Fisher Scientific) on N25 of H4 (via an H4(N25C) mutation). Nucleosomes for the comparison of H3-acetylation mimics against wild-type H3-containing nucleosomes were labelled on H3(R2C) instead of H4(N25C). Initial rates for each concentration of INO80 were plotted and analysed in GraphPad Prism 6.0 f with an

‘allosteric sigmoidal’ model; Hill coefficients were determined manually through a log-conversion of the data.

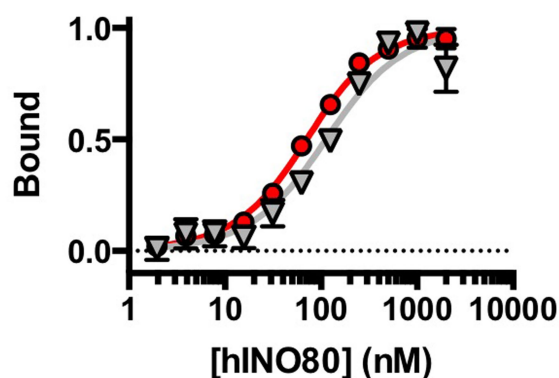
Nucleosome stability assays. Salt-stability assays were carried out on centrally positioned nucleosomes with Cy5 and Cy3 fluorescent labels on opposite ends of flanking DNA. Stocks of nucleosomes with wild-type or mutant histones were mixed with increasing concentrations of KCl and aliquoted into a 384-well microtitre plate. The intensity of the Cy3 donor label was then measured across different KCl concentrations, with higher intensity corresponding to decreased quenching and therefore unwrapping of the DNA tails from the nucleosome core.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Data have been deposited in worldwide Protein Data Bank (wwPDB); in the Electron Microscopy Data Bank with accession code EMDB 3954 for the INO80 core–nucleosome complex map, and in the RCSB Protein Data Bank with accession code 6ETX for protein coordinates. All other data are available from the corresponding authors upon reasonable request.

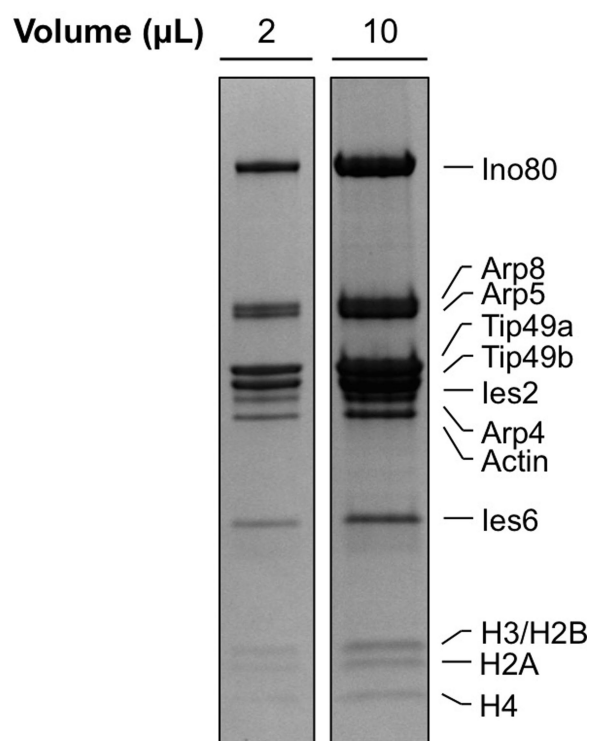
38. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
39. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
40. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
41. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
43. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
46. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
47. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.* **37**, D387–D392 (2009).
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
49. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

a



| | Ligand | $k_{1/2}$ (nM) | h |
|---|----------------------|-----------------|---------------|
| ▽ | Apo | 93.0 ± 11.0 | 1.6 ± 0.1 |
| ● | ADP.BeF ₃ | 51.0 ± 2.9 | 1.6 ± 0.2 |

b



c

```

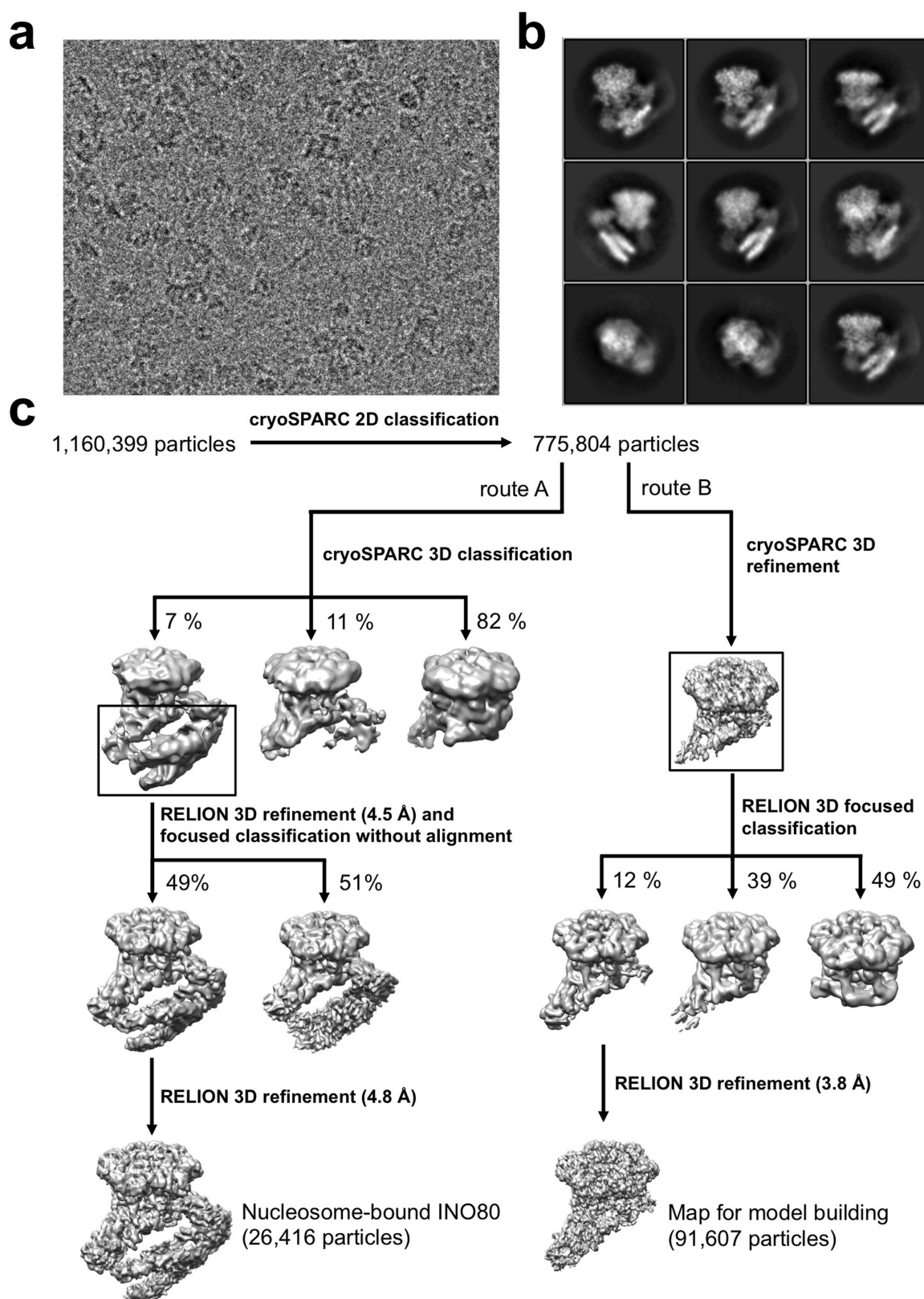
1 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACCCCGAGTA 47
48 CATGCACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGT 94
95 AATCCCCTTGGCGGTTAAACGCGGGGGACAGCGCGTACGTGCGTTT 141
142 AAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCG 188
189 GGATTGTCCAGGGCGGCCGCGGATGCATTAAATGCAGa t t 227

```

Extended Data Fig. 1 | Analysis of INO80–nucleosome complex sample.

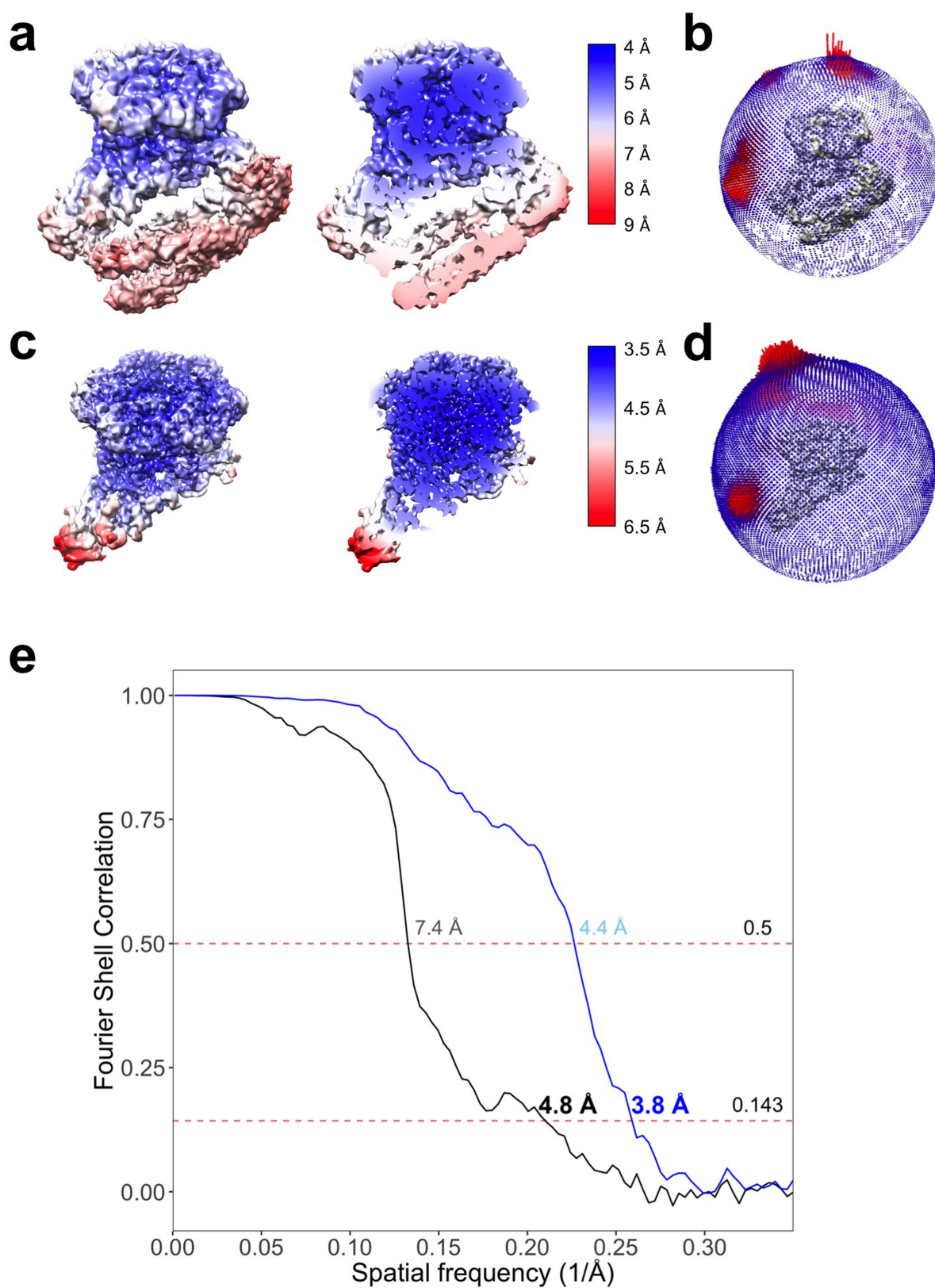
a, MST experiment of INO80 with nucleosome flanked by 60 and 12 base pair overhangs (60N12) (± 3 mM ADP•BeF₃). Raw data (top) were processed to analyse binding and cooperativity (bottom). Data points represent mean values with s.d.; $n = 3$ experimentally independent replicates. **b**, Gel of electron microscopy sample (INO80 + nucleosome). Two loadings are shown to enable assessment of INO80 stoichiometry

(left) or histones (right). $n = 3$ independent experimental measurements. **c**, DNA sequence of the 50N25 nucleosome used for the structure determination. The Widom sequence (yellow) is flanked by 50 base pairs on one side and 25 base pairs on the other. A three-base single strand overhang that remained from the restriction cleavage site is depicted in lowercase. For gel source data, see Supplementary Fig. 1.



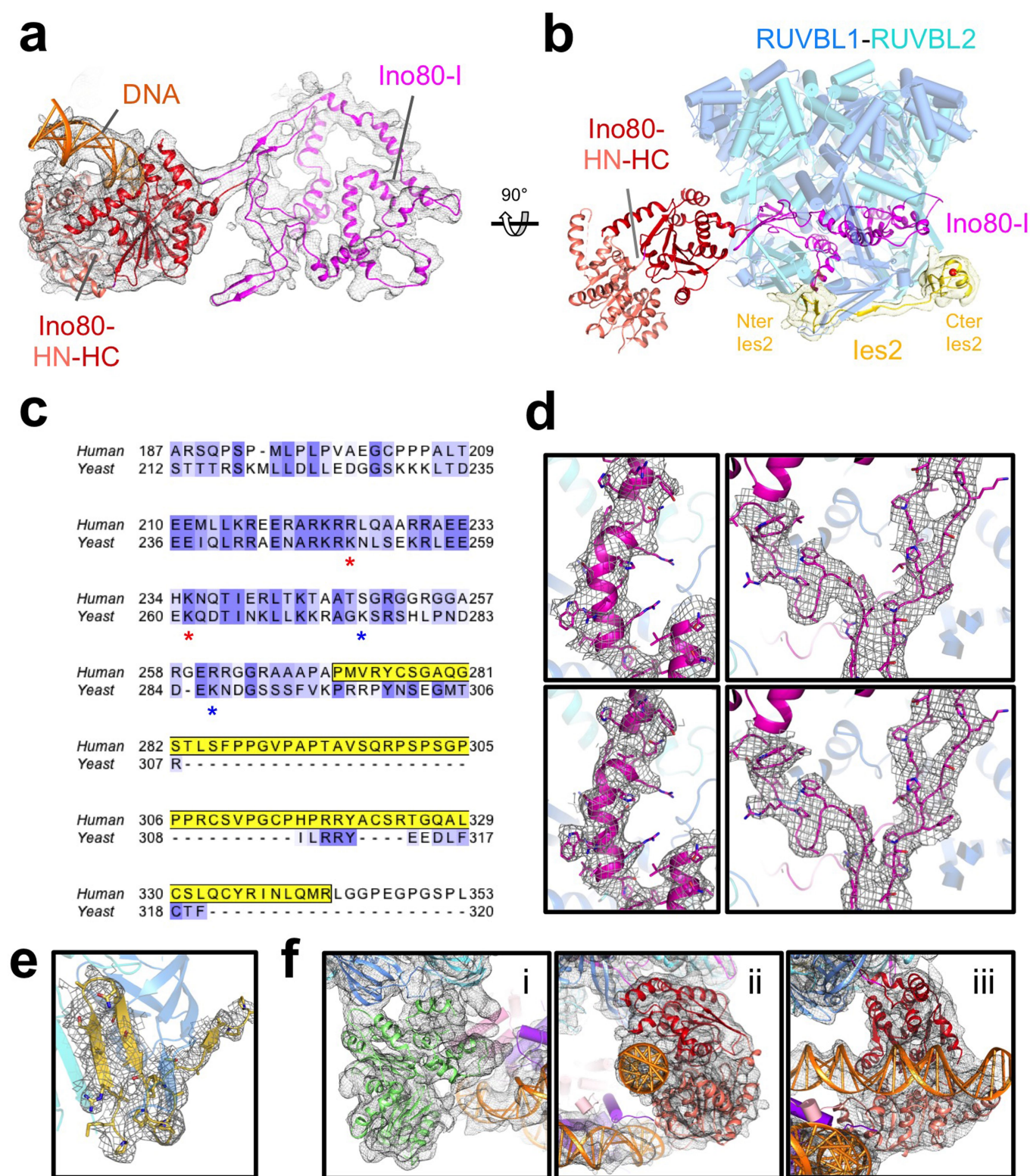
Extended Data Fig. 2 | Cryo-electron microscopy data processing of INO80–nucleosome complex. **a**, A typical micrograph out of the 5,479 micrographs generated. **b**, Representative 2D classes (from 100 generated)

obtained with RELION from 775,804 particles. **c**, Image processing scheme. Data were processed by two parallel pathways to obtain maps for model building.



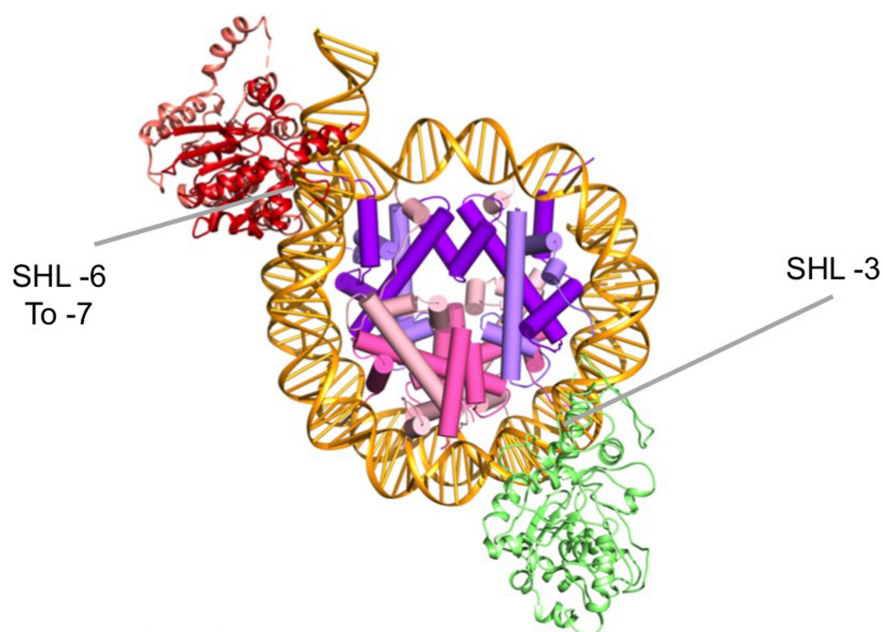
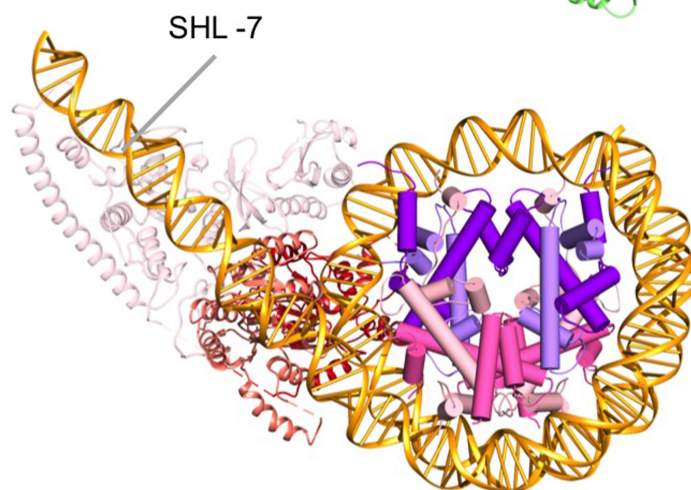
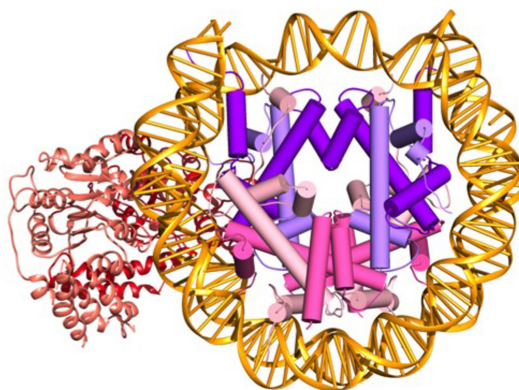
Extended Data Fig. 3 | Quality of the structures. **a**, Local resolution map of the INO80-nucleosome complex (4.8 Å) (left) and cut away (right). **b**, Angular distribution of these particles. **c**, Local resolution map of the

INO80 complex (3.8 Å) (left) and cut away (right). **d**, Angular distribution of these particles. **e**, Corrected FSC curves of the reconstructions.



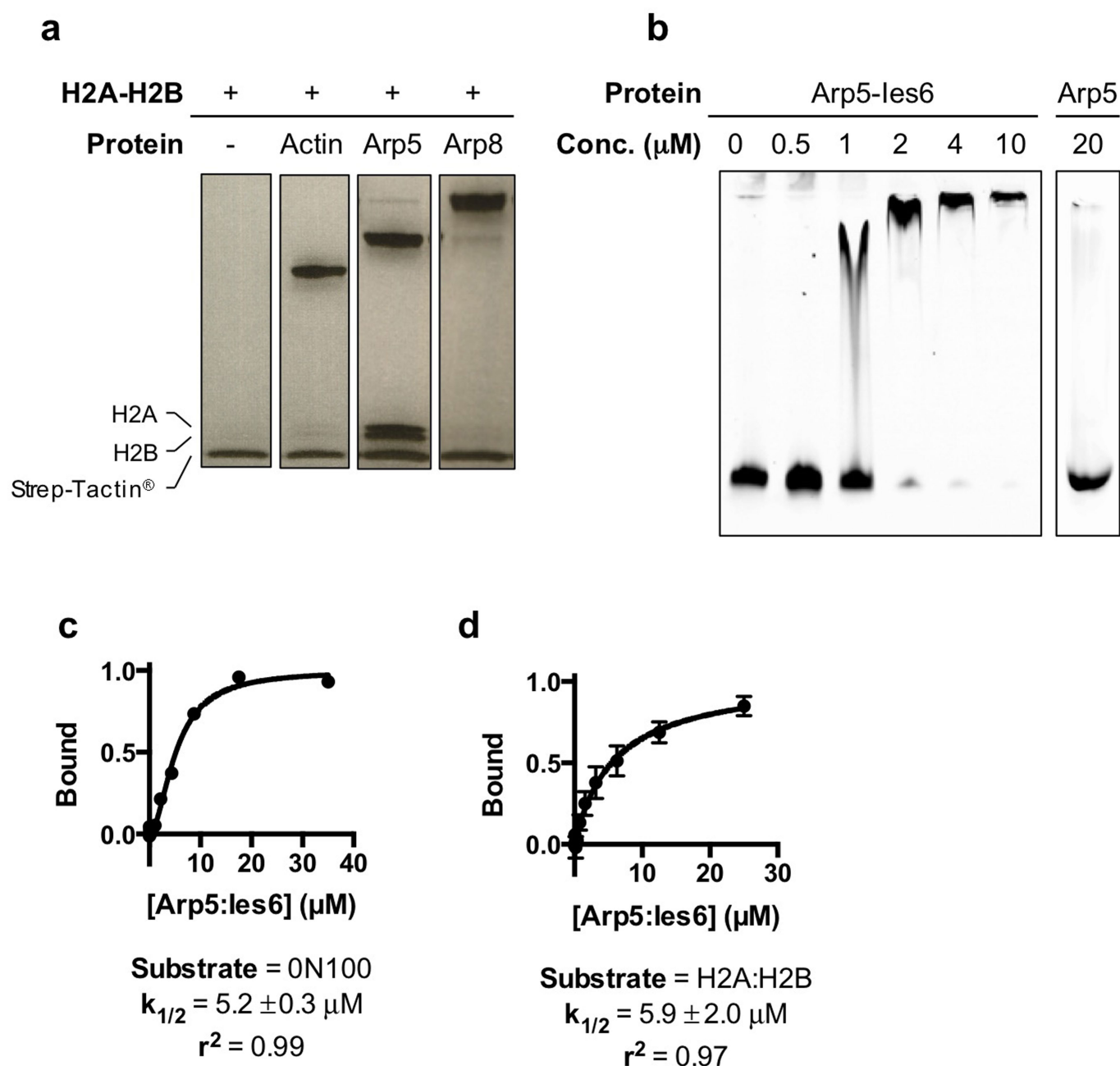
Extended Data Fig. 4 | Assessment of various structural features in the INO80-nucleosome complex. **a**, Overall fold of the INO80-I and motor domains. **b**, Locations of the INO80-I, motor domains and IES2 regions relative to the RUVBL1-RUVBL2 hexamer. **c**, Sequence alignment of the C-terminal regions of human and yeast IES2. The built part of the human IES2 structure is indicated by a yellow bar. Asterisks indicate lysine residues in yeast Ies2 that crosslink to Ino80-HN (red) or Ino80-HC (blue). **d**, Representative density from two regions of the INO80. Insert:

top, Density in the deposited 4.8 Å INO80-nucleosome map; bottom, improvement in density in the 3.8 Å map, which facilitated model building. **e**, Coordinates of IES2 showing formation of β -sheet secondary structure with RUVBL1 (chain E) and RUVBL2 (chain D) within the 3.8 Å map. **f**, Left, fit of ARP5 into 4.8 Å map (left). Centre, DNA and motor domains fit into the 4.8 Å map (centre). Right, perpendicular view of centre panel that shows the DNA crossing the motor domains.

INO80**CHD1****SNF2 (SHL2)**

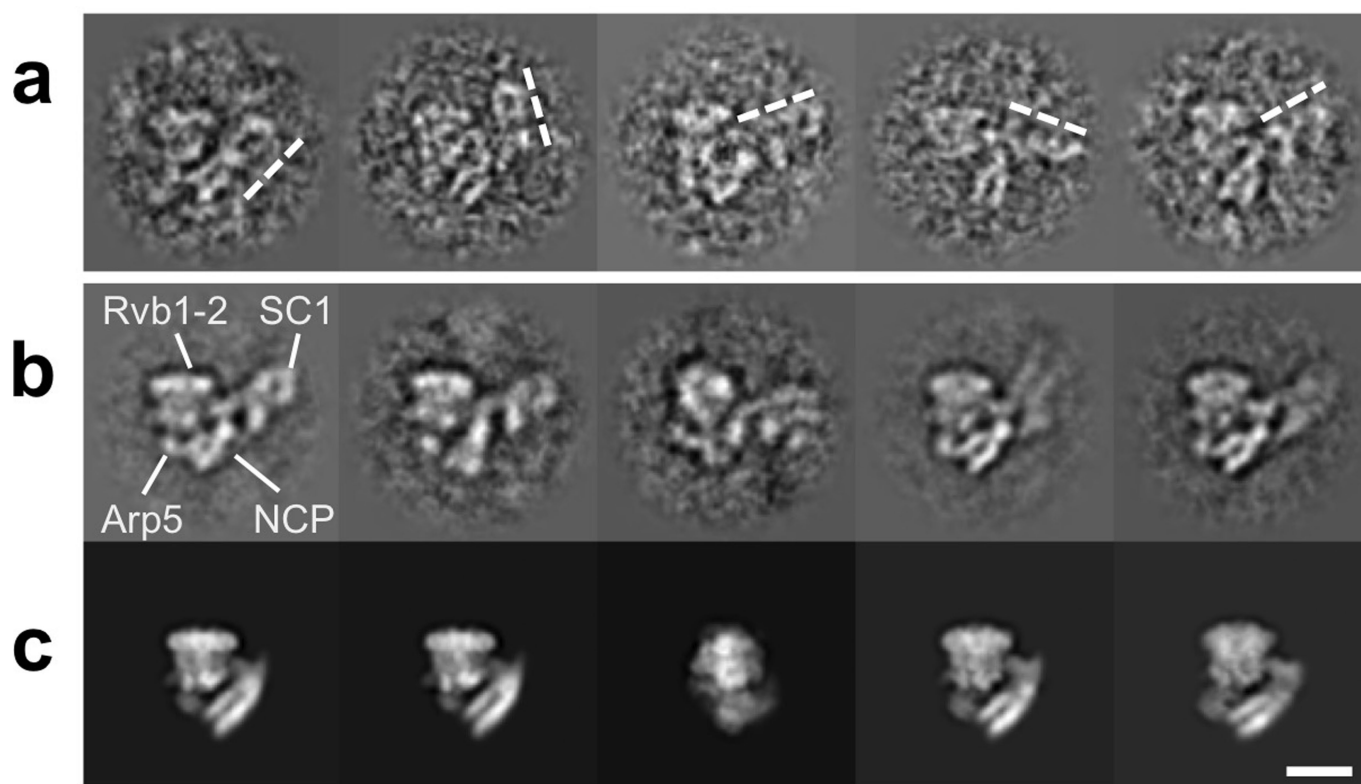
Extended Data Fig. 5 | Comparisons of INO80–nucleosome interactions with those of Chd1 and Snf2. Images are viewed from the top of nucleosome, showing that all the motor domains are located on one side and that ARP5–IES6 (green) contacts the other side of the DNA wrap. Chd1 induces an unwrapping of the DNA at the SHL –7 position owing in a large part to interactions with the accessory SANT and SLIDE domains.

Despite this unwrapping, the histone core remains largely unaltered. Although the Snf2–nucleosome structure does not induce unwrapping of DNA, it is only a fragment of the motor subunit and also lacks other accessory subunits of the SWI–SNF complex and so probably presents an incomplete picture of interactions or DNA distortions within the nucleosome in the complex.



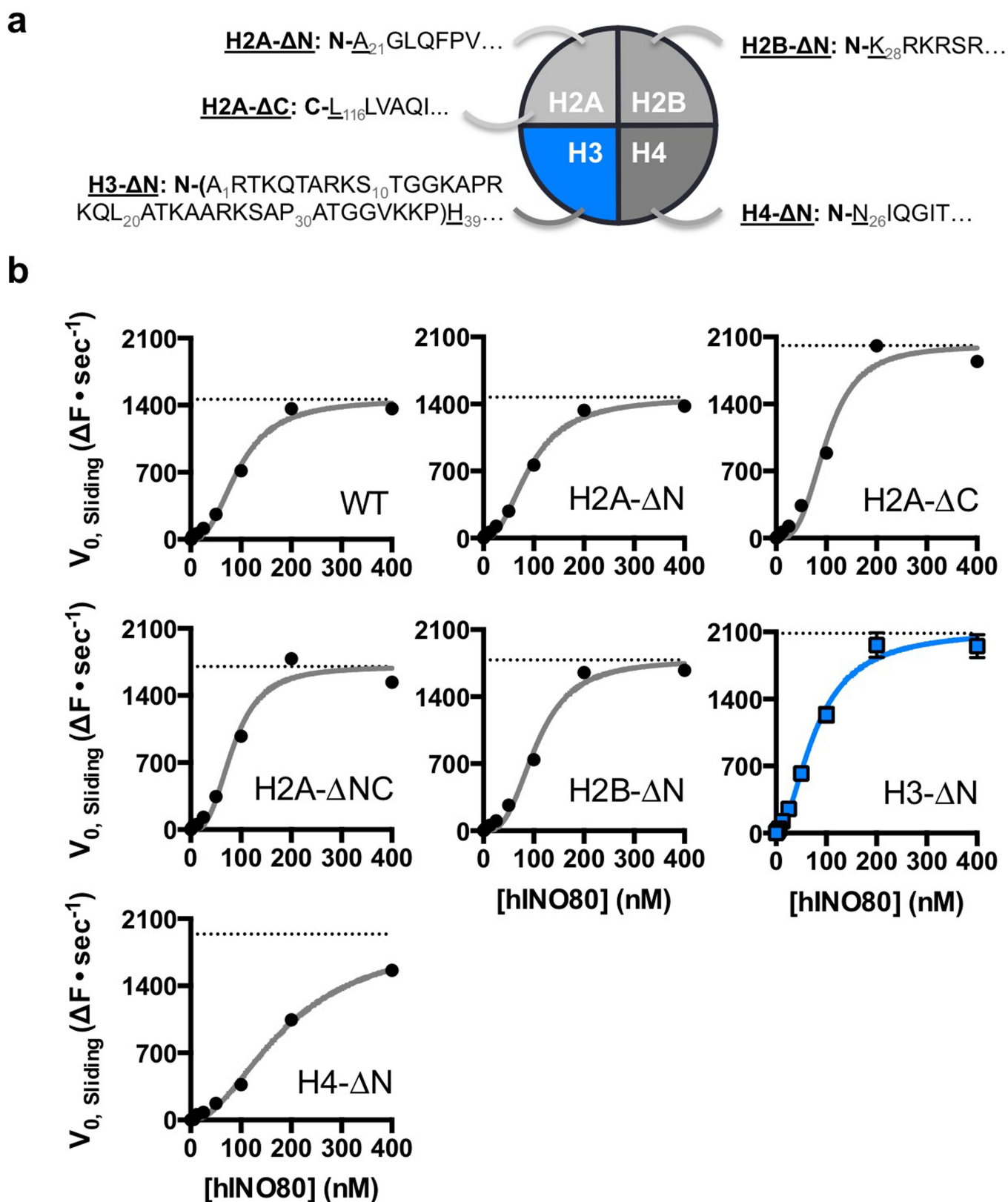
Extended Data Fig. 6 | Interaction of human actin, ARP5 and ARP8 with human H2A–H2B dimers assessed by in vitro pulldown. **a**, Actin and actin-related proteins were all expressed with a C-terminal double-Strep tag and used as bait to capture untagged H2A–H2B dimers. The result supports the position of ARP5 in the reported structure. Assay products were visualized by SDS–PAGE and Coomassie staining. $n = 1$. **b**, A comparison of ARP5–IES6 and ARP5 nucleosome-binding activity assayed by electrophoretic mobility shift assay, which demonstrates a lack

of nucleosome-binding activity by ARP5 at in vivo relevant concentrations in the absence of IES6. Nucleosomes were labelled with Alexa Fluor 488. Reaction species were visualized by fluorescent scan. $n = 1$. **c**, ARP5–IES6 and 0N100 nucleosome interaction measured by MST. **d**, ARP5–IES6 and H2A–H2B interaction measured by MST. For gel source data, see Supplementary Fig. 1. $n = 2$ biologically independent experiments in all the graphs. Error bars represent s.d. from the mean values.



Extended Data Fig. 7 | INO80 SC1 is flexible in the INO80–nucleosome complex. **a**, Individual particles (selected out of 775,804 particles in total) with RUVBL1–RUVBL2 oriented similarly, to show different orientations of SC1 (dashed lines). **b**, Two-dimensional class averages (~30 particles

each) showing different orientations of SC1 relative to RUVBL1–RUVBL2. **c**, Projections of the 3D reconstruction along the same angles of those in **b**, confirming the extra density as SC1. Scale bar, 100 Å.



Extended Data Fig. 8 | INO80 is regulated by H3 tails. a, Schematic of histone tail truncations used in this study. **b,** Initial nucleosome sliding rates of human nucleosomes that lacked different histone tails. Plots of raw data for each histone tail deletion, with V_{max} obtained after fitting the data

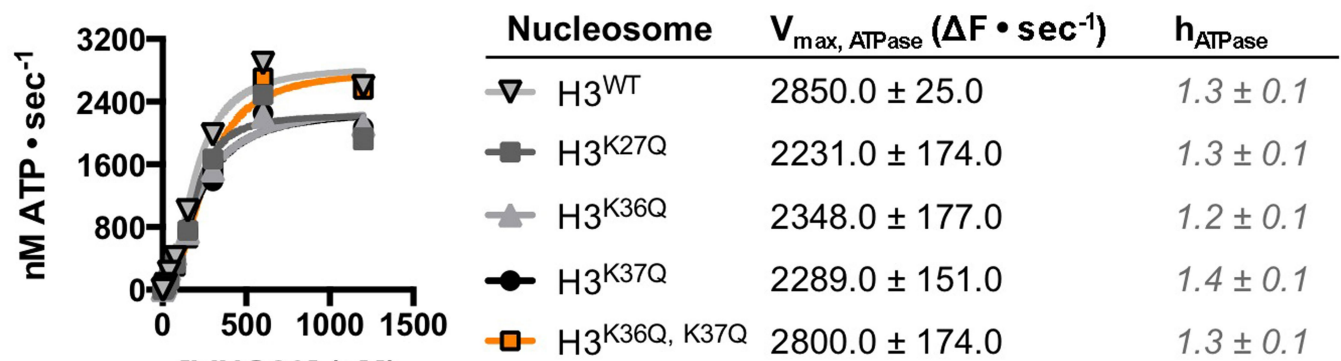
shown as a dotted line. Data are summarized in Fig. 4a. $n = 2$ biologically independent experiments in all the graphs. Error bars represent s.d. from the mean values.

a

| Nucleosome | V_{\max} (nM ATP • sec ⁻¹) | h_{ATPase} |
|----------------------------|--|---------------------|
| ON100 (H3 ^{FL}) | 2430.0 ± 69.0 | <i>1.4 ± 0.1</i> |
| ON100 (H3 ^{L20}) | 2050.0 ± 82.0 | <i>1.3 ± 0.1</i> |
| ON100 (H3 ^{P30}) | 2660.0 ± 92.0 | <i>1.4 ± 0.1</i> |
| ON100 (H3 ^{H39}) | 3150.0 ± 160.0 | <i>1.3 ± 0.1</i> |

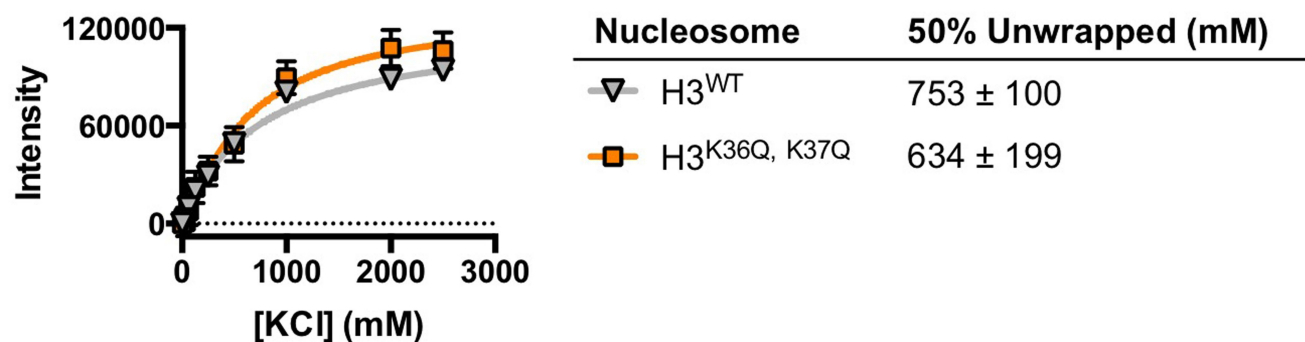
***Bold** = $h > 1.5$, *Italics* = $h < 1.5$

b



***Bold** = $h > 1.5$, *Italics* = $h < 1.5$

c



Extended Data Fig. 9 | INO80 is regulated by H3 tails. **a**, ATPase data and Hill coefficients for data shown in Fig. 4c. **b**, ATPase rates for mutations of the H3 tails. **c**, Nucleosomes carrying wild-type or mutated H3 tails show similar salt stability, which indicates that the mutations

have not altered the stability of nucleosomes. $n = 2$ biologically independent experiments in all the panels. Error bars represent s.d. from the mean values.

Extended Data Table 1 | Electron microscopy data collection, image processing and model refinement statistics

| | INO80 nucleosome complex (EMDB-3954) (PDB 6ETX) |
|--|---|
| Data collection and processing | |
| Magnification | 129,000 |
| Voltage (kV) | 300 |
| Electron exposure (e-/Å ²) | 80 |
| Defocus range (µm) | -2.0 to -4.0 |
| Pixel size (Å) | 1.13 |
| Symmetry imposed | C1 |
| Initial particle images (no.) | 775,804 |
| Final particle images (no.) | 26,416 |
| Map resolution (Å) | 4.8 |
| FSC threshold | 0.143 |
| Map resolution range (Å) | 4.0 to 8.0 |
| Refinement | |
| Initial model used (PDB code) | 2YQQ, 3LZ0, 5AV9, 5O9G, 50AF |
| Model resolution (Å) | 4.8 |
| Map sharpening <i>B</i> factor (Å ²) | -100 |
| Refinement (Phenix) | |
| Map correlation coefficient (whole unit cell) | 0.88 |
| Map correlation coefficient (around atoms) | 0.74 |
| Model composition | |
| Non-hydrogen atoms | 38,759 |
| Protein residues | 4,753 |
| Nucleic acid residues | 300 |
| Ligands | 7 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.002 |
| Bond angles (°) | 0.447 |
| Validation | |
| MolProbity score | 1.58 (93 rd percentile*(0Å - 99Å)) |
| Clashscore | 8.4 |
| Poor rotamers (%) | 1.5 |
| Ramachandran plot | |
| Favored (%) | 98.2 |
| Allowed (%) | 1.75 |
| Disallowed (%) | 0.02 |

CAREERS

SPACE PARITY Germany is launching its first female astronauts **p.399**

CANADA Diversity efforts in academia are falling short **p.399**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

ADAPTED FROM GETTY



COLUMN

Use video to cut through jargon

Films can take science to a crucial but often-overlooked audience, says **Adrian A. Smith**.

Three years ago, I realized that I had never given my mum a clear explanation of what I do as a research scientist. I knew I needed to change that.

I was considering how I communicated my science to non-scientists, and began to see my mother as a part of an audience I hadn't fully reached. She knew the generalities of my work — that I study ants' behaviour and had published many peer-reviewed papers about it — but she'd never read any of those studies. And, given that I'd never walked her through the specifics of any of them, none was truly accessible to her.

I decided that I could solve that problem — for my mum, and for the many other people who don't read primary scientific literature, perhaps because of a paywall or because of the unfamiliar technical format and language. I

wanted her and others to be able to learn about my research.

When all this occurred to me, it was November 2015 and I had just started my first faculty position, at the North Carolina Museum of Natural Sciences (NCMNS) in Raleigh, where I head a research lab. My mother, Cindy, was coming to visit for the US Thanksgiving holiday. So, the day after the holiday, I asked her to come to work with me. I had decided to conduct a long-overdue experiment in science communication: it was time to sit down and talk to Mum about the specifics of my research.

We went to the museum's studio. There, I would film our unscripted chat for a video to accompany my institution's press release about an upcoming paper of mine (A. A. Smith *et al.* *J. Exp. Biol.* **219**, 419–430; 2016). With a few

visual aids — including pictures of the ants that I study and diagrams of the chemicals that they use to communicate with each other — and with three cameras pointing at us, I started explaining my results on fertility and sexual dimorphism in cuticular hydrocarbon profiles of three species of trap-jaw ant (*Odonotomachus*). Or, in less-jargonized language, how ants use a unique chemical language to communicate about sex and fertility.

For 40 minutes, the cameras rolled and I stumbled through descriptions of my latest findings, as my mother gave me live feedback on what she did and didn't understand. I asked whether she remembered what kinds of chemicals I studied, and handed her a diagram. She grimaced. She had heard me talk about cuticular hydrocarbons, but for her to understand their importance, I needed to show her that ►

► they were chemicals, and explain how ants used them to communicate with each other.

After our video session, I spent a week editing what we'd filmed into a four-minute video summary of the paper. The press release describes the research in the standard way — in third person, with me as the lead researcher quoted in the middle. However, embedded in the document is the video, which intersperses our dialogue with stills of the ant images and diagrams that I used.

I think that the clip captures both my research and a slice of my relationship with my mum that in turn helps to make the science more engaging and relatable: even as I'm trying to present her with a summary of my work, she's cheekily interjecting one-liners about how worker ants and queens are just like sons and mothers. To understand the 'what' and 'why' of my research, one could read the press release — or, just watch our video.

The video ends with me asking my mother why she thinks that my research is important, after hearing about the study. "It's part of our world," she says. "We need to understand what it does. How we can get along with it better." At that point, I felt confident that my mum and I were on the same page in terms of why I was doing this work.

Today, when I ask my mum what she says when her friends ask about what I do for work, she has a succinct answer. "I tell them you research ants," she says. "Then I say that we made a video that explains the details and why you do what you do."

JARGON BE GONE

This was the third video I'd made as a way of translating primary scientific research for a non-scientist audience. A year or so earlier, I'd realized that most people would be lost in my papers' technical jargon and formatting. I'd also noticed that when family members and friends sent me articles about new research, they weren't providing links to a journal's table of contents or PDFs of a manuscript — they were sending popular-media news stories about the work.

I came to understand that if I wanted my science to find its way to the same sources that my family and friends were using, I would need to rethink my publication process. As a scientist with an interest in digital media, I had a direct path for getting first-person narratives about my work to a global mass-media audience.

Today, most media outlets source their science news from institutional press releases announcing new discoveries. These write-ups often appear on news aggregators, such as the American Association for the Advancement of Science's 'EurekAlert!'. Reporters cover science news by including in their stories perspectives

and quotes from their own sources, beyond the information in an institutional release. But some sites simply repost press releases along with stills and videos. This new, more-direct intersection between scientists, aggregators and science-news consumers is where I found my path to the public.



Adrian Smith and his mother, Cindy, prepare for their live science chat.

In the past three years, I have produced and posted ten videos tied to institutional press releases about scientific research papers originating from my or my colleagues' labs. We want to use press releases as a way to make our research narratives directly accessible to a mass-media news audience — just as I made my paper accessible to my mum.

Here's how it works: we include a URL in the press release that leads reporters to a related YouTube video. We post these press releases on EurekAlert! with an embargoed period of two or three days before publication. US and international news outlets, including *Wired's* UK edition, *The Washington Post* and *der Standard*, a daily Austrian newspaper, have picked up five of the ten press releases as the basis for their own stories, and they have added our videos alongside their written coverage. Cumulative views on the YouTube videos picked up by those outlets range from 5,000 to about 62,000, whereas views on videos associated with releases that did not get major news coverage range from 1,000 to 3,900.

When these videos were released, the number of subscribers to my personal YouTube page was a mere smattering, about 200 or 300, compared with the number of views for each of my clips. Clearly, the much larger number of

viewers compared to subscribers was a direct result of media interest in the press releases. And these view counts are a conservative measure of engagement: they include YouTube views only, and not those, for example, from instances in which news outlets, such as the *Washington Post* or *National Geographic News*, requested the original video and posted it directly on their sites.

Working with public-information officers to distribute media-rich press releases has given my colleagues and me the ability to present the value we see in our work to a science-curious public ourselves. We have been able to reach much larger audiences than we could have by simply publishing in journals. The engagement numbers are persuasive, even when these papers are published through open access. One of the papers we promoted was published in *PLoS ONE*, where page-view numbers are public (F. J. Larabee and A. V. Suarez *PLoS ONE* **10**, e0124871; 2015). Since the paper's release in May 2015, it has accumulated around 12,200 page views — but the video about the paper has received more than 62,300 views.

First-person accounts of science were not a part of my life when I was younger. I am a first-generation university graduate with no immediate or extended family members who are involved in scientific careers. As a child, I'd never known a working scientist. When I was filming that video with my mum, I realized that I was presenting myself as a professional scientist to a family member who also had never had a personal connection to science before me. Making videos and using press releases to distribute them has helped me to introduce myself and my colleagues to the world as scientists. I now view the impact of my research not just in relation to the metrics around my journal articles, but also in terms of how well I can make my work available to those outside my profession.

Online science videos and the press-release distribution system allow for direct access by and dialogue between researchers and science-news consumers. Adding first-person narratives and reshaping science-news information is not impossible for scientists who are willing to communicate their research actively. By making this content about our science readily available to any viewer, we can reach people who are interested in science but can't read original manuscripts in a journal for whatever reason.

If you don't believe me, just ask my mum. ■

Adrian A. Smith is head of the Evolutionary Biology & Behaviour Research Lab at the North Carolina Museum of Natural Sciences and a research assistant professor in biology at North Carolina State University in Raleigh.

ADRIAN SMITH

"I now view the impact of my research in terms of how well I can make my work available to those outside my profession."

TURNING POINT

Space pioneer

MARKUS GLOGER/DIE ASTRONAUTIN

Insa Thiele-Eich, a meteorologist and scientific coordinator at the Meteorological Institute of the University of Bonn, is one of two women training for the opportunity to become Germany's first female astronaut.

How did you get involved in the astronaut training programme?

I applied for Die Astronautin, a private initiative that is independent of the European Space Agency (ESA). It is aiming to put the first German woman into space, and is eager to highlight female role models in science for young girls. It got a lot of publicity in March 2016 when only women were invited to apply. More than 400 candidates sent in CVs, academic grades, reference letters and short videos describing their goals.

How were the candidates narrowed down?

Through an interview, an extensive questionnaire about lifestyle, exercise habits and medical history, and medical, psychological and other tests. Nicola Baumann, a mechanical engineer and fighter pilot, and I were selected in April 2017 to start astronaut training. Nicola has since left the competition and astrophysicist Suzanna Randall has been chosen as her replacement.

How did it feel to be chosen?

It's hard to describe. After such a long, intense time together, we were all very close. Honestly, it took me a couple of days to be able to smile about it.

What does your training involve?

In August, we started to practice putting on space suits and drinking water in a zero-gravity environment, as part of aircraft flights that simulate weightlessness at Roscosmos, the Russian space agency. It's like taking a ride in a centrifuge. I'm also obtaining my pilot's licence and learning how to scuba dive.

When do you expect to go to space?

We'll find out a year before the actual flight. The mission is currently scheduled for 2020, which, in the field of space flight, could mean anything from 2020 to 2024. Successful launches of crewed craft by SpaceX and Boeing are crucial for the private sector and will give us options for flying to the International Space Station. The companies' launches have been moved to 2019, which has pushed back our potential launches. Until then, Suzanna and I train part-time. I'm not in a rush. I like my job, coordinating research among 90 scientists at an interdisciplinary research centre.



Has training altered your research interests?

I have put more focus on remote sensing, which is the use of images collected from satellites or high-flying aircraft. I hope to determine how remote-sensing data could benefit public-health studies on the ground.

What were your initial career goals?

As the daughter of an astronaut in Houston, Texas, I always knew that I wanted to be a scientist, and I was interested in the interdisciplinary research being done on the space station. I studied meteorology at the University of Bonn, but in Germany, there are almost no tenured positions any more. I was on contracts that had to be renewed regularly so I always wanted an alternative to deal with the insecurity. Being an astronaut was my plan B.

Why wasn't going to space your plan A?

There weren't many opportunities through which to become an astronaut — especially when no German woman had ever been one. I couldn't apply to NASA because I'm not a US citizen. And the last ESA application campaign was in 2008, and there are no plans for another for at least three or four years.

Has Die Astronautin invigorated your efforts to attract women to science?

Yes. I did a TEDx talk on a favourite topic of mine — how to get women into space. When I was selected last April, I had a rude awakening. In my daily job, I don't really perceive gender differences or pay much attention to the fact I'm female. But I received many sexist remarks from the public and some interviewers. A lot of work must be done for society to be more open and treat women equally. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for clarity and length.

WORKFORCE

Canadian diversity

The Canadian academic workforce is not as diverse as the nation's general labour force, according to a report by the Canadian Association of University Teachers in Ottawa, which represents about 70,000 academic professionals. The report, called *Underrepresented and Underpaid* (see go.nature.com/2jay1d2), finds that 1.4% of university professors identify as Aboriginal, compared with 3.8% of the nation's labour force. Black people make up 2% of university teachers but 3.1% of the nation's workforce. And although 48.5% of all assistant professors are women, just 27% of full professors are, the report found. Women are also underrepresented in science, technology, maths and engineering across Canadian institutions, the report said. They comprise 24.8% of all full-time faculty members in physical and life sciences and technologies; 27.6% in agriculture, natural resources and conservation; and 20.6% in maths, computer technology and information sciences. Female full-time faculty members earn an average of Can\$123,225 (US\$97,700) — 90% of their male counterparts' pay.

POSTDOCS

Grant success

Postdoctoral researchers in biomedical fields who win a particular grant from the US National Institutes of Health (NIH) are more likely to receive subsequent grants from the NIH than are other applicants, a study finds. Published online by the US National Bureau of Economic Research, *The Impact of Postdoctoral Fellowships on a Future Independent Career in Federally Funded Biomedical Research* reviewed NIH grant records for individuals who applied for the Ruth L. Kirschstein National Research Service Award 'F32' Individual Postdoctoral Fellowship between 1996 and 2008. Lead author Misty Heggeness and her co-authors then examined those researchers' NIH application and funding patterns up to 2015. Overall, 18.3% of F32 applicants went on to receive NIH research awards, with 13.3% winning an NIH 'R01' project grant, which supports independent research. Those who received an F32 award were one-half to two-thirds more likely to receive subsequent NIH research awards. They were also less likely to be black or Asian, or to be older than 37. The success rate for F32 applicants was 24.9% in 2017, according to NIH data.

WASTELAND OF SAND AND ICE

Rapid response.

BY TOMÁS MCMAHON

Autonomy is a beautiful thing.

06:00: The time of rising. Not the best time to learn of an asteroid hurtling past Uranus at more than 40 million metres per second.

06:30: The time of bathing. AbyssRho's computer monitors were awash with triangulation data and extrapolated vectors. The results were in, the likelihood of terrestrial impact looking shockingly like the percentage of bacteria a good hand soap claimed to kill.

07:00: The time of breaking fast, the time to break old rivalries. One of the benefits of the Deep Space Identification Network being a multinational project was that NASA and Roscosmos received exactly the same data. An agreement was made and the asteroid named: 2037 KD.

The Post-War Asian Union wires the money and takes a step back from whatever AbyssRho does with it. Far simpler for both parties that way.

08:30: The time for toiling, preparing for the inevitable. Missile codes were requested, authorized thrice and ultimately given. PerUN, GUATAUVA, Lei Gong, turned away from their endless showdown amid their ruined brethren to gaze up into space. The plan was simple: if KD was to reach to Earth, it would have to dance through hell.

On any other day, three major superpowers relinquishing possession of their military satellites' kinetic warheads would have been seen as a sign of imminent world peace. That day, however, was not like any other.

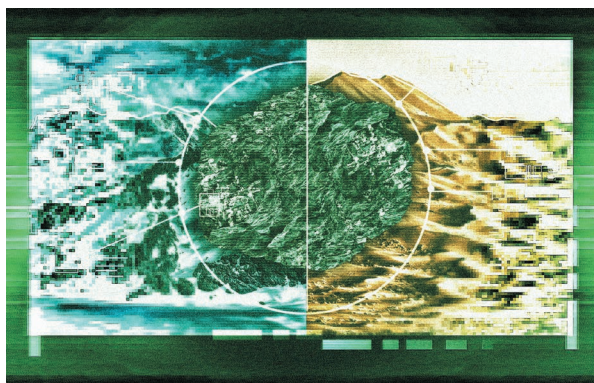
Meanwhile, near the Bay of Bengal, unnoticed, a hurricane began to build.

14:00: The time to return to work. It was around Saturn that KD had been imaged, distant and low-resolution, initially appearing almost as an anticlimax. The asteroid wielded no scythe nor rode an ashen horse, however, upon closer inspection, its true deadliness soon became apparent. A smaller package simply meant a smaller mass and, in turn, a smaller likelihood of interception.

17:30: The time KD passed Jupiter. About the time the button was pressed. By the

number of warheads up in the air, anyone could have thought that a nuclear apocalypse had somehow snuck up on them.

Their city of smoke pillars rapidly dissipated by the winds of the encroaching storm, many missiles wouldn't make it past the vast swathes of debris in Earth's orbit. Those that did would bear all of humanity's hopes of preservation. It was just a good thing they were machine and not mortal.



GUATAUVA automatically receiving data from the Deep Space Identification Network, no input required, certainly saves NASA plenty of precious time. Far simpler for both parties that way.

22:55: An insignificant time. 2037 KD's sudden deceleration came as a surprise. Dodged by the bullet is an uncommon phrase for obvious reasons, but here it was most appropriate. How else would one describe a body decelerating from one-sixth the speed of light to one-forty-millionth?

23:00: The time of rest. The Torino level slashed, just like most of AbyssRho's funding. *The price to pay for wasting the Union's warheads.*

Adrift within the spacecraft graveyard of Earth's orbit, its solar-sail finally fully collapsed, KD 2037 gazed down towards the blanketed planet beneath.

00:00: The time of sleeping soundly. Dashed upon the Himalayas, the storm smothered all of central Asia in an anarchic maelstrom. Above the chaos, KD fragmented, two, smaller objects splitting off. Within the opaque clouds, they fell gracefully, explosively stabilized with fire and wind-torn parachute. Like marionettes on unseen strings, the two were dragged off course, plummeting into sand and ice.

The Deep Space Identification Network reports potential asteroid threats and justifies PerUN's possession of kinetic warheads, in turn making it technically not a military satellite. Far simpler for both parties that way.

In the predawn hours of the following morning, the hurricane had torn itself apart. After KD flew into a debris cloud and did not exit on the other side, it was assumed that it had suffered a similar fate.

As the last wisps of the ultimately short-lived storm spiralled beneath, a surveillance satellite, one of the last few left in operation, imaged an abnormality in the Siberian permafrost. Shallow but nevertheless noticeable, a crater, far north of Lake Cheko, thin tracks leading off to the southwest and, at their end, 2037 KD-01.

03:00: The time for toiling had come early. When the recovery teams at last cleared the snow aside, approximately half the fragment's mass was found at the site of its initial impact, what crawled out having only recently stilled, hull shattered and frozen.

Less than two hours later, at the bottom of a furrow dug into the leading edge of a sand dune, the second fallen object finally revealed its metallic self. In comparison with its twin, 2037 KD-02 had enjoyed a soft landing, the Gobi Desert acting to cushion its fall, allowing it to remain in operation for several hours — at least until the dune swallowed the second of the two rovers.

Autonomy truly is a beautiful thing.

06:00: The time of rising, tension, suspicions and, most of all, a search for any sort of explanation. Disassembly and reassembly on behalf of the Union and Russia, those actually in possession of a fragment, while NASA, through gritted teeth, was already planning a mission to salvage what it could of the orbiting body. As for missions themselves, the probes' had certainly ended in failure. The planet was clearly a wasteland of sand and ice, covered by an atmosphere of opaque gases and extreme winds, 2037 KD transmitting back as follows: *No life on Earth.* ■

Tomás McMahon is an A-level student in England. Until located, it can be presumed he is both sketching and out cycling in the Surrey hills at the same time.

ILLUSTRATION BY JACEY